



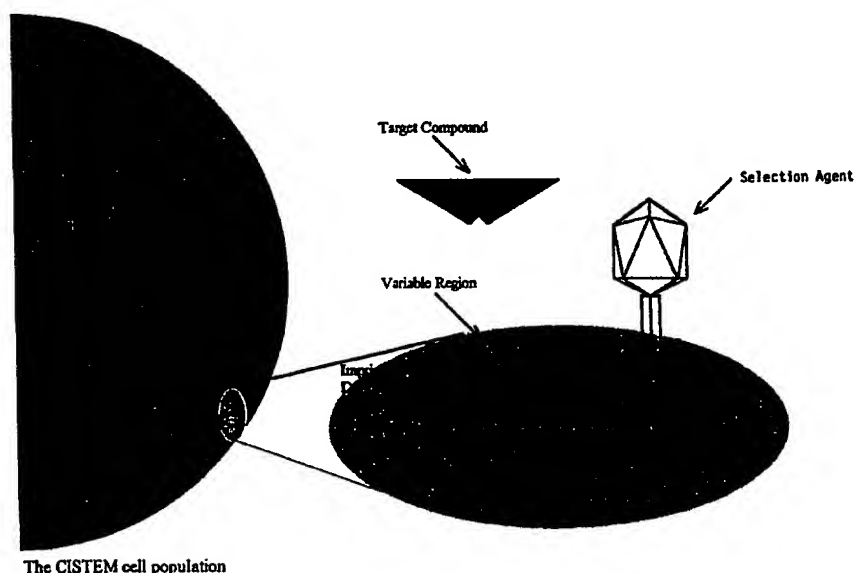
PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
 International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

| | | | |
|---|--|--|--|
| (51) International Patent Classification ⁶ : G01N 33/48, 33/50, 33/569, 33/68, C12N 11/16 | | A1 | (11) International Publication Number: WO 99/30151 |
| | | | (43) International Publication Date: 17 June 1999 (17.06.99) |
| (21) International Application Number: PCT/EP98/07922 | | (81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). | |
| (22) International Filing Date: 7 December 1998 (07.12.98) | | <p>Published</p> <p><i>With international search report.</i></p> <p><i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p> | |
| (30) Priority Data: 97121451.5 5 December 1997 (05.12.97) EP 98890217.7 23 July 1998 (23.07.98) EP | | | |
| (71) Applicant (for all designated States except US): INTERCELL BIOMEDIZINISCHE FORSCHUNGS- UND ENTWICKLUNGS GMBH [AT/AT]; Hockegasse 77, A-1180 Vienna (AT). | | | |
| (72) Inventors; and (75) Inventors/Applicants (for US only): VON GABAIN, Alexander [AT/AT]; Hockegasse 77, A-1180 Vienna (AT). HIRSH, Aaron [US/US]; 2035 Amherst Street, Palo Alto, CA 94306 (US). | | | |
| (74) Agents: SONN, Helmut et al.; Riemergasse 14, A-1010 Vienna (AT). | | | |

(54) Title: METHOD AND CELLS TO SCREEN COMBINATORIAL LIBRARIES



(57) Abstract

This invention is a method for identifying and isolating at least one of the following: a polypeptide of interest and/or a gene thereof; a target compound and/or a gene thereof, wherein the polypeptide of interest, or the target compound, or both are unknown. The method is based on the steps of: (1) displaying a library of a polypeptide of interest on the surface of a population of prokaryotic or eukaryotic cells; (2) exposing the population to a target compound; and (3) exposing the population to a selection agent. Interaction among the three components "imprints" the polypeptide: that is, the interaction selects for those cells in which the polypeptide interacts with the target compound in a prescribed manner.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | | | |
|----|--------------------------|----|--|----|--|----|--------------------------|
| AL | Albania | ES | Spain | LS | Lesotho | SI | Slovenia |
| AM | Armenia | FI | Finland | LT | Lithuania | SK | Slovakia |
| AT | Austria | FR | France | LU | Luxembourg | SN | Senegal |
| AU | Australia | GA | Gabon | LV | Latvia | SZ | Swaziland |
| AZ | Azerbaijan | GB | United Kingdom | MC | Monaco | TD | Chad |
| BA | Bosnia and Herzegovina | GE | Georgia | MD | Republic of Moldova | TG | Togo |
| BB | Barbados | GH | Ghana | MG | Madagascar | TJ | Tajikistan |
| BE | Belgium | GN | Guinea | MK | The former Yugoslav Republic of Macedonia | TM | Turkmenistan |
| BF | Burkina Faso | GR | Greece | ML | Mali | TR | Turkey |
| BG | Bulgaria | HU | Hungary | MN | Mongolia | TT | Trinidad and Tobago |
| BJ | Benin | IE | Ireland | MR | Mauritania | UA | Ukraine |
| BR | Brazil | IL | Israel | MW | Malawi | UG | Uganda |
| BY | Belarus | IS | Iceland | MX | Mexico | US | United States of America |
| CA | Canada | IT | Italy | NE | Niger | UZ | Uzbekistan |
| CF | Central African Republic | JP | Japan | NL | Netherlands | VN | Viet Nam |
| CG | Congo | KE | Kenya | NO | Norway | YU | Yugoslavia |
| CH | Switzerland | KG | Kyrgyzstan | NZ | New Zealand | ZW | Zimbabwe |
| CI | Côte d'Ivoire | KP | Democratic People's Republic of Korea | PL | Poland | | |
| CM | Cameroon | KR | Republic of Korea | PT | Portugal | | |
| CN | China | KZ | Kazakstan | RO | Romania | | |
| CU | Cuba | LC | Saint Lucia | RU | Russian Federation | | |
| CZ | Czech Republic | LI | Liechtenstein | SD | Sudan | | |
| DE | Germany | LK | Sri Lanka | SE | Sweden | | |
| DK | Denmark | LR | Liberia | SG | Singapore | | |
| EE | Estonia | | | | | | |

- 1 -

METHOD AND CELLS TO SCREEN COMBINATORIAL LIBRARIES

Combinatorial chemistry (reviewed in Janda, 1994), in which a diverse library of laboratory-constructed molecules is screened for structures that exhibit properties of interest, is a relatively novel and potentially powerful technology, promising useful application in three general areas. First, combinatorial chemistry may be used to identify molecules with highly specific adhesive or enzymatic activities useful in biomolecular research, clinical therapy, or diagnostics. Second, by screening a library for molecules that interact specifically with a known compound of significant interest, combinatorial chemistry may assist in identification of the known compound's native interactants (e.g. Phizicky et al., 1995). Third, combinatorial chemistry may be used for epitope mapping, serving to elucidate the surface of structural interaction between two molecules of interest (e.g. Phizicky et al., 1995).

The current combinatorial chemistry of proteins, powerful though it may be, has shown two distinct limitations. In many libraries, the diverse proteins are presented on a platform that fails to confer much structural integrity; as a result, the proteins can be too plastic to exhibit a stable shape and consistent properties. Equally as important, the screening procedure can be labor-intensive, and factors outside of a protein's exhibition of the properties of interest often influence its isolation by a screen. This occurs in most screens reported in the literature (e.g. Phizicky et al., 1995). A related technology, two-hybrid systems, has proven useful in the latter two applications mentioned above, namely, interaction studies and epitope mapping (Phizicky et al., 1995). However, two hybrid systems place a very strict limit on the kinds of interaction that may be studied: both components of the interaction must be proteins, and they must be readily transported into the nucleus.

It was an object of the present invention to provide a method of polypeptide combinatorial chemistry and polypeptide interaction identification that resolves or circumnavigates

CONFIRMATION COPY

- 2 -

many of the difficulties encountered by other technologies. (The term polypeptide is used herein to comprise all types of peptides or proteins, including glycoproteins, proteoglycans, lipoproteins, acetylated proteins or (other) processed peptide or protein forms). It was a further object to provide a fast and reliable screening method for combinatorial or other peptide libraries, which is also versatile and broadly applicable.

These objectives are met by a method for identifying and isolating at least one of the following: a polypeptide of interest, a target compound or a gene of a polypeptide of interest or a gene of a target compound, wherein at least one of said polypeptide of interest or said target compound is unknown comprising the steps

- a) constructing a cell population by engineering it to express a membrane polypeptide which has said polypeptide of interest in its extra- or intracellular domain, said extra- or intracellular domain being exhibited in one or more copies on each cell's surface directly or indirectly interacting with a selection site specific for a selection agent, such that upon binding of said target compound to said polypeptide of interest said selection site of cells which exhibit such a binding of said target compound to said polypeptide of interest is modified in a manner that said selection agent may be used for selecting those cells which exhibit said binding of said target compound to said polypeptide of interest from cells which do not exhibit said binding,
- b) exposing said cell population to said target compound or a population of said target compounds for a period of time sufficient for said target compound to bind to said polypeptide of interest,
- c) exposing said cell population simultaneously with or after step b) to said selection agent for a period of time sufficient for said agent to interact with said selection site,
- d) selecting for clones which show said binding of said target compound to said polypeptide of interest ,

- 3 -

- e) optionally allowing said selected clones to multiply and
- f) isolating said polypeptide of interest and/or said target compound and/or said gene of said polypeptide of interest and/or said gene of said target compound.

In this invention, a library of polypeptides is exhibited in one or more extra- or intracellular domains of a membrane polypeptide expressed by a cell population. Selective pressures exerted by selection agents eliminate those cells whose polypeptide sequences fail to bind to a preselected target compound. Thus, from a cell population that may exhibit a diversity of polypeptide sequences, cells whose sequences bind the target compound are selected and e.g. amplified by growth. The term "polypeptide imprinting" is used to refer to the use of the method according to the present invention to select cells exhibiting polypeptides that bind the target compound or exhibit specific properties.

With this invention, a wide variety of prokaryotic and eukaryotic polypeptides may be imprinted, serving e.g. creation or discovery of polypeptide sequences with high affinity for a predetermined target, interaction studies, or epitope mapping. In more sophisticated embodiments of the invention, polypeptides may be imprinted to exhibit the following specific properties: multivalent binding to a predetermined target; high affinity for a target and low affinity for one or more molecules closely related to the target; inhibitory activity in a biological pathway (antagonism); excitatory activity in a biological pathway (agonism); or enzymatic activity.

It is also possible to screen not only for a polypeptide library but also coincidentally or alternatively, for a population of target molecules.

As mentioned above, the present invention may be performed by displaying the polypeptide to be imprinted (the polypeptide of interest) in the extra- or intracellular domain of a membrane protein.

Preferred embodiments of the present invention are described hereinafter and in the examples and figures.

- 4 -

This invention is the product of the confluence of three distinct areas of research: i. the study of prokaryotic and eukaryotic cellular membranes and their complex constituents; ii. the study of the evolution of microbial populations under the selective pressures of toxins, bacteriocins, or viral parasites; and iii. the design and screening of polypeptide combinatorial libraries.

With the present invention, a polypeptide with promising properties can immediately be identified, amplified and isolated, making the present method extremely sensitive. It is possible to select one polypeptide that binds to a ligand out of a background of 10^6 non specific polypeptides as a minimum, but in there is no evidence for an upper limit on sensitivity regarding the present method per se.

With the present method all cells of the combinatorial library of polypeptides and/or target compounds are screened at once, rather than serially, providing results in far less time than many alternative technologies. The present invention is easy to use, because the important parameters are easily adjustable for the man skilled in the art. Once a set of cells is selected, the genes for the polypeptides or the target compounds (if these are also expressed by the cells) are easily isolated.

The polypeptide of interest may be displayed by any of the wide ranges of cell types known and used for displaying foreign proteins in extra- or intracellular domains (e.e. on the cell surface). This allows the present method to avoid the biases of any one expression system. The polypeptide is preferably expressed in a membrane protein which is biochemically appropriate for the target compound or for the system wherein the isolated polypeptide/target pair is intended to be applied (e.g. as a vaccine or as a therapeutic).

Combinatorial chemistry, which is one of the major components of the present invention, consists of two or three steps, often performed in iterative cycles. The first step is the generation of a diverse population of molecules, the so-called combinatorial library. In the second step, the molecules in the combinatorial library are partitioned

- 5 -

according to their exhibition of properties sought by researchers, such as affinity for a particular compound or enzymatic activity. In a third step that is practiced in some but not all embodiments of combinatorial chemistry, the molecules found to exhibit the desired properties are replicated or resynthesized. The cycle may then be repeated, beginning either at the first step, with the introduction of new variation to the molecular population, or at the second step, with the partitioning of the population according to the exhibition of desired properties. When a combinatorial approach is successful, the cycle of generating molecular diversity, whittling down diversity in favor of the few molecules that display the desired properties, and, in some cases, amplifying the numbers of these molecules, ultimately yields molecules exhibiting the desired specific activities..

The immune system is essentially a genetically coded combinatorial library, complete with steps of variation, selection, and replication: A vast diversity of antibodies is displayed on the surface of B-cells (variation), and specific recognition of antigen by any given cell triggers that cell's proliferation (partitioning and replication). It is not surprising, then, that the first in vitro genetically-coded combinatorial libraries were directly appropriated from the immune system: B-cells were immortalized through hybridization to mouse myeloma lines (Marks et al., 1991). Later, in order to circumnavigate difficulties of hybridization, and to utilize a greater part of the antibody repertoire, *E. coli* were engineered to express and secrete the variable chains of antibodies. However, because secreted antibodies were "untagged" (Janda, 1994)-viz. not physically linked to their genetic codes-screens for antibodies that bound antigen had to be performed according to techniques that allowed researchers to trace from each selected antibody to the gene that encoded it. This severely limited the number of clones that could feasibly be screened. The advent of phage-display, or "fusion-phage", libraries made it possible to screen variable proteins that were "tagged" (Janda, 1994), and thus competent for immediate replication.

- 6 -

Combinatorial libraries of fusion-phage are composed of filamentous bacteriophage or phagemid that have been engineered to bear semi-random exogenous proteins as sandwich fusions near the end of the adsorption protein pIII (Janda, 1994, Marks et al., 1991). The method of screening these libraries for phage that exhibit affinity for a target compound is known as "affinity purification": the target is immobilized on a solid surface or column, over which the phage library is washed; phage bound to the target are eluted, amplified by growth in bacterial culture, and then subjected to yet another round of partitioning. This cycle is usually iterated until a tractable number of clones are isolated. Phage libraries have been used for epitope mapping (Phizicky et al., 1995)-viz. pinpointing the region of a larger protein that is involved in binding a target compound of interest-as well as for deriving novel ligands to targets. Entire single-chain variable-fragment antibodies (scFv's) have been inserted into phagemid pIII (Marks et al., 1991), and libraries of scFv's or semi-random oligopeptides have been screened for binding to a variety of medically significant polypeptides, nucleic acids, and molecules not previously known to bind proteins.

However, while fusion phage technology has been responsible for significant strides forward in combinatorial chemistry - such as the recognition that "tagged" libraries would permit screening of a vastly increased number of sequences - phage libraries themselves have not escaped several besetting difficulties. First, in affinity purification of phage libraries, factors besides affinity for the target compound contribute significantly to each clone's selection. For example, the collection of clones selected from a single library by affinity purification can vary dramatically with the technique used for eluting phage from the solid substrate of target compound (Phizicky et al., 1995). Second, phage adsorption protein, pIII, may fail to provide exhibited proteins with a sufficiently stable platform. Many polypeptides acquire structural rigidity only in certain constraining conformations; otherwise they remain extremely plastic. Therefore, to be functionally active

- 7 -

structures, many polypeptides will require a foundation other than pIII.

A promising solution to problems with the platform for polypeptide display was introduced when a combinatorial library of proteins was displayed in an extracellular loop of a bacterial outer membrane protein (Brown, 1992). To construct this library, between 1 and 20 semi-random oligonucleotides, each one 33 bases long, were inserted into the genetic site encoding an extracellular loop of the outer membrane porin LamB. Like many membrane proteins that have been the subjects of topological (Boyd, 1994) and antigen presentation (Georgiou et al., 1997) studies, LamB intercalates the outer membrane even when exogenous polypeptide sequences have been inserted into one or more of the protein's extracellular domains. Thus, when *lamB* genes bearing insertions were used to transform *E. coli*, each transformant expressed LamB in which one extracellular domain exhibited a semi-random, exogenous oligopeptide between 11 and 220 amino acids long.

To screen the library for proteins exhibiting affinity for the target compound iron oxide, the *E. coli* population was placed in a suspension of iron oxide particles, binding was allowed to occur, and a magnet was used to recover bacteria that had adhered to metal (Brown, 1992). The selected portion of the population was amplified by bacterial growth, and the partitioning procedure was repeated. After four cycles of selection and amplification, fresh variation was introduced according to a strategy that would not erase sequence-information accrued thus far. Another four cycles of selection and amplification then yielded four clones, three of which bore a consensus sequence for binding iron oxide.

While the combinatorial library in LamB introduced the promising possibility of using bacterial outer membrane proteins as platforms for presenting libraries of diverse polypeptides, the selection procedure applied to the library was labor-intensive and very limited in its applicability. The target compound, iron oxide, was conveniently chosen to allow screening for magnetism - a procedure that could not be applied to screening for the majority of interesting interactions that polypeptides exhibit. The present invention

- 8 -

offers a user-friendly, efficient, and precise method of selection, capable of screening libraries of all sorts of polypeptides presented on the platform of a membrane polypeptide.

The membranes bounding any prokaryotic or eukaryotic cell, referred to here as the outer or inner membrane, present an extremely complex surface to the extra- or intracellular environment. These molecules - referred to here as membrane polypeptides - exhibit a vast range of distinctive conformational, chemical, and electrical properties. Surveyed elsewhere in greater detail (Singer, 1990), the membrane polypeptides are here categorized into several structural families. The conformational and functional differences between these families are relevant to this invention because membrane polypeptides provide the underlying structures that are used to display exogenous sequences, and are selected for interaction with the target compound. Moreover, this invention often capitalizes on the intrinsic properties of membrane polypeptides by imprinting them to exhibit properties or perform functions that are similar to their natural ones. Finally, the deliniation of structural families presented here should also provide a profile of the vast range of properties exhibited by polypeptides on the cell membranes, for it is in part this range that makes the capabilities of this invention so broad.

Single-pass transmembrane polypeptides (reviewed in Singer, 1990), also called bitopic polypeptides, have one intra- and one extra-cellular domain, and a midsection that passes through the membrane. The terminal domains fold into tertiary structures, while the transmembrane section often forms a hydrophobic α -helix, in which no more than 2 of the 20 transmembrane residues may be ionic. Singer (Singer, 1990) divides bitopic polypeptides into types Ia, Ib, and II (figure 1). Types Ia and Ib exhibit the amino terminus on the extracellular side of the outer membrane. Distinguishing type Ia is an amino terminal signal sequence that directs export to the membrane, and is cleaved before final membrane insertion. Type Ia polypeptides are ubiquitous, and in eukaryotes they are the predominant type. Because the transmembrane sequence

- 9 -

of type Ia polypeptides is typically closer to the carboxyl terminus, the extracellular domain is often large. Consequently, type Ia polypeptides are competent for functions involving high-affinity adhesion or highly specific binding to especially extracellular targets. Among type Ia polypeptides are receptors for growth factors and for peptide hormones; immune system antigen recognition molecules; and molecules responsible for cell adhesion (Singer, 1990, Staunton et al., 1990). Proclivity for adhesion or highly specific binding to extracellular targets make type Ia polypeptides suitable for use in many embodiments of this invention, especially those in which a polypeptide is imprinted to bind to an extracellular target.

Type Ib polypeptides do not have a cleavable signal sequence. The transmembrane stretch is typically close to the amino terminus, and therefore there are few extracellular residues (Singer, 1990). This makes type Ib polypeptides suitable for sophisticated cytoplasmic functions, and interactions, but poorly suited for interactions with targets in the extracellular medium.

Type II bitopic polypeptides are oriented with their carboxyl terminus on the exterior side of the membrane (figure 1). As in type Ia bitopic polypeptides, the bulk of type II molecules is usually on the exterior of the cell. Type II polypeptides are typically enzymes with extracellular active sites, such as glycosyl transferases, or highly specific extracellular receptors, such as the asialoglycoprotein receptor (Singer, 1990). Competence for extracellular enzymatic or receptor activity makes type II bitopic polypeptides suitable for use in certain embodiments of this invention, in which extracellular domains are imprinted to bind or enzymatically modify targets in the extracellular medium.

Multipass transmembrane polypeptides (reviewed in Singer, 1990), also called polytopic polypeptides, traverse the lipid bilayer in two to fifteen or more passes. Like thread stitched into cloth, the polypeptide begins on one side of the membrane, passes through it, loops into the space on the other side of the membrane, then once again traverses the membrane

- 10 -

to expose either its end or yet another loop leading to more transmembrane passes (figure 1). Although referred to and often illustrated as "a loop", each exposed region of a multipass polypeptide exhibits not only a unique sequence of amino acids, but also a distinctive tertiary structure. Like single-pass transmembrane polypeptides, multipass ones are suspended in the lipid bilayer by the hydrophobicity of transmembrane strands. However, in multipass polypeptides, these strands do not invariably twist into α -helices, because when multiple strands neighbor one another, they can adopt an alternative strategy for hydrogen bonding between peptide bonds: if each strand adopts the secondary structure of a β -strand, these strands can curl together into a closed cylinder, allowing hydrogen bonds to form between the peptide bonds on separate strands. The resultant tertiary structure is called a β -barrel (figure 1). It is extremely stable, and can be released from the outer membrane without denaturing tertiary structure. In several preferred embodiment of this invention, the extremely stable, polytopic β -barrel OmpA is imprinted to bind to various targets.

Polytopic polypeptides that traverse the membrane in α -helical conformation often function as specific receptors for messengers or hormones, such as yeast mating pheromone and many others. These and other α -helical polytopic receptors are categorized by Singer (Singer, 1990) as type III polypeptides. Their capacity to bind target molecules specifically, and sometimes multivalently, makes α -helical polytopic receptors ideal for use in many embodiments of this invention. Moreover, selection systems according to the present invention are easy to construct when the membrane polypeptides involved in the selection are polytopic.

Many prokaryotic and eukaryotic cell-surface polypeptides are fixed to the lipid bilayer not by a hydrophobic transmembrane strand, but rather by a glycolipid anchor (reviewed in Cross, 1990) (figure 1). An anchor consists of a lipid molecule, such as dimyristyl glycerol or alkylacyl glycerol, adorned by a multi-residue sugar complex, which is, in turn, covalently linked to the membrane polypeptide. Among eukaryotes, two signaling elements of a protein direct the

- 11 -

addition of a glycolipid anchor: a well-defined amino terminal signal sequence, which enables transport into the lumen of the endoplasmic reticulum, and a carboxy terminal glycolipid-addition sequence, which is not precisely defined but generally consists of a short hydrophobic domain located 10-12 residues carboxy terminal to a pair of small residues (Cross, 1990). Once appended to the carboxy terminus of a protein, the glycolipid anchor itself can direct polypeptide insertion in a targeted region of the outer membrane.

Lacking an intracellular domain, glycolipid-anchored outer membrane proteins cannot perform cytosolic or periplasmic functions. Instead, they provide protection from the extracellular environment, or function as enzymes or adhesins recognizing exclusively extracellular substrates. At least six leukocyte surface proteins, many other eukaryotic proteins belonging to the immunoglobulin superfamily (Cross, 1990), and the trypanosome variable surface glycoprotein (the protective protein that the trypanosome sequentially alters to avoid detection by the host immune system) are anchored by glycolipids. These examples illustrate the characteristics of glycolipid-anchored proteins that are important to their use in embodiments of this invention: capacity for highly specific adhesion and plasticity that allows variation of the anchored protein without disturbing a robust system for export to and insertion in the outer membrane.

Among eukaryotic cells, the diverse array of surface-exposed proteins is adorned by yet another tier of complex molecules with distinctive properties, namely, the saccharides (reviewed in Lis et al., 1993). The oligosaccharides adorning glycoproteins generally consist of less than 15 residues. However, owing to the diversity of residues and covalent linkages that permit assembly into branched as well as linear structures, oligosaccharides exhibit thousands of distinct structures. The polysaccharides of integral membrane proteoglycans are larger than oligosaccharides, and can contain dozens of residues.

The oligo- or polysaccharide molecule attached to a particular site in a polypeptide is not a simple function of the local amino acid sequence. The same polypeptide sequence

- 12 -

receives distinct saccharide chains when expressed in different cells, and separate copies of a polypeptide expressed in a single cell type can have different sugars attached at a given site. Because sugar residues are sequentially appended, each by a particular enzyme to which the protein is exposed in the course of the processes of translation, folding, and export to the outer membrane, the glycosylation at a site is largely a product of the extent and duration of the site's exposure to each enzyme. Exposure of a site is, in turn, affected by the way a protein folds and the rate at which it moves through the steps from translation to insertion. Consequently, the glycosylation at a site can be altered in unpredictable ways by changes to the local sequence or to the sequence elsewhere in the polypeptide (Lis et al., 1993).

Diversity of molecular structure and relatively complete exposure to the extracellular environment make the saccharides appended to proteins effective receptors of specific substrates outside the cell. Glycoproteins and proteoglycans are involved in the intercellular recognition and adhesion underpinning sperm-egg interactions, blood clotting, and inflammation (Lis et al., 1993, Singer, 1990). The roles they play, via adhesion, in these important processes make the glycoproteins and proteoglycans ideal for use in certain embodiments of this invention.

Therefore, in a preferred embodiment of the present invention, the membrane protein is an α -helical polytopic receptor, a glycolipid-anchored polypeptide, a polytopic integral membrane polypeptide, a membrane-bound isoform of a secreted polypeptide, a bitopic integral membrane polypeptide, a glycoprotein or a proteoglycan.

It is significant to some embodiments of this invention that insertion of polypeptides into the membrane and translocation of secreted polypeptides across the membrane are performed by the same cellular machinery (Singer, 1990). Among eukaryotes, integral membrane polypeptides inserted into the endoplasmic reticulum (ER) have the same cleavable or noncleavable amino terminal signal sequences as completely translocated polypeptides. This is consistent with the fact

- 13 -

that insertion and secretion depend on the same cytoplasmic signal recognition protein (SRP), which transfers both inserted and secreted polypeptides to the SRP receptor on the ER membrane. Similarly, among bacteria, membrane insertion and translocation seem to require the same gene products, namely secA and secY (Singer, 1990). Among prokaryotes and eukaryotes alike, there are many examples of secreted and membrane-bound polypeptide isoforms. Often the only difference between the isoforms is the existence of a hydrophobic stretch of residues or a glycolipid attachment site near the carboxy terminus of the membrane-bound polypeptide (Singer, 1990).

The process of type III polypeptide intercalation into the membrane also has implications that are significant to this invention. The cellular machinery that inserts bitopic polypeptides and translocates secreted ones is also responsible for intercalating α -helical multipass polypeptides into the outer membrane. This is clear from studies that show that i) type III polypeptides sometimes have the same signal sequences as bitopic and secreted polypeptides, and ii) membrane intercalation of type III polypeptides requires the same gene products as insertion of bitopic polypeptides. Several experimental approaches, including gene fusion, have shown that the cellular machinery for inserting polypeptides in the outer membrane intercalates most type III polypeptides by translocating one hydrophilic extracellular loop at a time. The polypeptide is thus intercalated as a series of independently inserted segments (Singer, 1990).

Most embodiments of this invention require a topological map of the imprinted polypeptide, and a technique for introducing new sequences to specified extracellular domains of the imprinted polypeptide. Both of these needs are met by gene fusion techniques.

A vast variety of polypeptides, especially the polypeptides reviewed above have already been topologically mapped using gene-fusion techniques. In these studies, the gene for a membrane polypeptide was joined to a gene coding for a "reporter", a protein that is functionally active only when it is located on a particular side of the cell membrane. This genetic construct coded for a hybrid polypeptide in which

- 14 -

the membrane polypeptide bore the reporter moiety as a mid-sequence insertion or as a C-terminal extension. Provided the foreign moiety disrupted neither sorting of the polypeptide to the appropriate membrane, nor insertion of the polypeptide into the membrane, the reporter was localized in the same way as the immediately adjacent region of the membrane polypeptide. This region's location on the exterior or interior of the membrane was then inferred from the reporter's activity, or lack thereof. The entire polypeptide's topology was deduced by assaying reporter activity in a set of hybrid polypeptides, each of which bore the reporter at a different place in the membrane polypeptide's sequence (these gene-fusion mapping techniques are reviewed in Boyd, 1994). Not only has this approach provided topological maps of most of the polypeptides commonly used in gene technology, especially the proteins described above; it can also be used to determine the topology of any membrane polypeptide that has not yet been mapped, but is a candidate for imprinting by this invention.

In addition to a topological map of the imprinted polypeptide, each embodiment of this invention also requires a technique for inserting exogenous peptide sequences into one or more extra- or intracellular sites in the imprinted polypeptide. Alternatively, exogenous peptide sequences may replace existing sections of the extra- or intracellular domains of membrane.

Because gene-fusion topology studies used a set of hybrid polypeptides, each bearing the reporter in a region whose location relative to the membrane was to be mapped, several techniques were developed to enable insertion of the reporter gene at precisely specified sites in the membrane protein gene. These techniques include the following: i. the use of vectors with restriction sites that occur at unique locations in the membrane protein gene; ii. an oligonucleotide-directed deletion mutagenesis method; and iii. polymerase chain reaction methods (Boyd, 1994). Each of these techniques can be used in this invention to insert exogenous sequences into a specified site in the imprinted polypeptide; an extension of these techniques, which is used in this invention to introduce

- 15 -

new variation to extra- or intracellular domains, is described below.

Besides providing topological maps and techniques for inserting exogenous peptide sequences, gene-fusion studies have collectively demonstrated that introduction of an exogenous sequence of peptides to an extra- or intracellular domain of a membrane polypeptide does not interfere with the sorting of the polypeptide to its appropriate membrane, or with the polypeptide's normal insertion into the membrane. This capacity of extra- or intracellular domains to accommodate foreign oligopeptides is critical to the construction of combinatorial libraries of exogenous oligopeptides in the extra- or intracellular domains of membrane polypeptides. This capacity is also, therefore, critical to certain embodiments of this invention. It has been observed in various bacterial, yeast, and mammalian polypeptides, including the examples described in the present specification.

An extracellular domain of lamB, an outer membrane porin of *E. coli*, has been shown to accept exogenous polypeptides up to 165 amino acids in length without disturbing normal sorting or insertion of the protein (Brown, 1992). Foreign sequences have also been inserted into extracellular domains of *E. coli* PhoE (phosphate-limitation-inducible outer membrane pore), and FhuA (ferrichrome uptake receptor), and the insertions had no effect on protein sorting or insertion (Georgiou et al., 1997). In the specific case of *E. coli* OmpA, one of the most abundant polypeptides of the bacterial outer membrane (Morona et al., 1985), oligonucleotides rich in restriction endonuclease sites were inserted into the *ompA* gene at sites corresponding to protein loops II and IV, and the altered protein was expressed and exhibited normally (Freudl, 1989). Subsequently, influenza haemagglutinin, beta-galactosidase, foot and mouth disease viral protein 1 (FMDV-VP1), malaria antigens, and several other antigenic polypeptides up to 56 amino acids in length were successfully displayed in OmpA loops III and IV (Georgiou et al., 1997). Furthermore, two lines of evidence demonstrated that these exogenous polypeptides adopted stable folded structures. First, *ompA* fusion proteins exhibiting viral

- 16 -

antigens were shown to elicit intense virus-specific immune response in rabbits (Georgiou et al., 1997). Second, Cex exoglucanase displayed on an extracellular domain of OmpA hydrolyzed cellulose as effectively as conventionally immobilized enzymes (Francisco et al., 1993).

In yeast, extracellular domains of a number of membrane proteins accept exogenous peptide sequences without exhibiting disturbance to normal sorting and insertion (Boyd, 1994, Boder and Wittrup, 1997). Higher eukaryotic membrane polypeptides have also exhibited the capacity to accept foreign inserts at extracellular domains. When a polypeptide is expressed in eukaryotic cells, gene fusion studies may insert extra glycosylation sites to function as reporters: only polypeptide domains on the luminal side of the ER are exposed to glycosylating enzymes, and luminal domains become extracellular domains. Extra glycosylation sites were inserted into extramembraneous regions of mouse muscle nicotinic acetylcholine receptor: the inserted sites did not disrupt normal sorting or insertion of the outer membrane polypeptide; and, sites inserted in extracellular receptor domains were in fact glycosylated (Chavez et al., 1991).

In summary, gene fusion mapping techniques contribute to this invention in four ways. First, these techniques have been widely used to determine the topology of countless outer membrane polypeptides, many of which are described above and may be imprinted by this invention. Second, these techniques may be used to provide topological maps for polypeptides that have not yet been mapped, but may be candidates for imprinting by this invention. Third, gene fusion techniques provide this invention with commonly practiced methods for inserting an exogenous peptide sequence into a specified region of a membrane polypeptide; the use of these techniques in the invention, as well as the straightforward extension that enables the insertion of variable sequences, is described below. Fourth, gene fusion topology studies have demonstrated that the introduction of an exogenous sequence to an extra- or intracellular domain of many mapped polypeptides does not interfere with the polypeptides' export to or insertion in the membrane.

- 17 -

Certain embodiments of this invention involve alteration of a polypeptide's mode of attachment to the outer or inner membrane. Because insertion into the membrane and translocation across the membrane are handled by the same cellular machinery, secreted, transmembrane, and glycolipid anchored isomorphs of a single polypeptide are often interconvertible. Moreover, secreted and membrane-bound isomorphs can use the same amino terminal signal sequence, so in many cases, the only alteration required for conversion of a polypeptide to its isomorph is a change at the polypeptide's carboxy terminal. Attachment of a 20 residue hydrophobic sequence to the carboxy terminal of a secreted polypeptide is often sufficient to construct an isomorphic polypeptide that is anchored by a transmembrane span; this technique for converting a secreted polypeptide into a membrane-bound one is described in the literature (Singer, 1990). Alternatively, transmembrane or secreted polypeptides can be converted to glycolipid anchored variants by addition of carboxy terminal signal sequences from the precursors of glycolipid anchored polypeptides; this technique is also described in the literature (Cross, 1990). In some cases, a glycolipid anchor not only attaches an exogenous polypeptide to a cell membrane, but also directs sorting of the anchored alien polypeptide. In such cases, a single glycolipid anchor sequence is all that is required to direct an exogenous polypeptide to the membrane, translocate it, and anchor it (Cross, 1990).

In certain embodiments of this invention, two or more outer or inner membrane polypeptides are fused to construct a chimeric multipass polypeptide comprising extra- and/or intracellular domains of both constituent polypeptides. This manipulation has been successfully performed with a variety of prokaryotic and eukaryotic polypeptides. For example, in *E. coli* ompF-ompC chimeric protein 3601 (Fourel et al., 1990), the region of ompF comprising extracellular loops 1 through 5 is fused to the region of ompC comprising extracellular loops 6 and 7. When expressed in *E. coli* SM1005 (F^- DlacU169 *rpsL relA thiA fibB gyrA ompC ompF*), this chimeric polypeptide is properly inserted in the outer membrane (Fourel et al., 1990). In a striking example of the construction of eukaryotic

- 18 -

chimeric polypeptides, novel transmembrane polypeptides were created as follows. A region of the human asialoglycoprotein receptor H1, which included the polypeptide's transmembrane domain (20 hydrophobic residues, 40-61 in the wild-type protein), was tandemly repeated two, three, or four times. In each case the H1 repeats were joined by relatively hydrophilic linkers, such as the hexapeptide HRPSLG. These constructs were normally inserted into membrane, and luminal domains were normally glycosylated.

The general principle that often makes fusion of membrane proteins possible has already been outlined: the above description of the process of membrane protein insertion emphasizes that α -helical transmembrane strands of polytopic membrane proteins are inserted sequentially and independently of one another. An implication of this mechanism of intercalation of transmembrane strands is that almost any chimeric polypeptide that is constructed of an amino-terminal signal sequence followed by multiple @20-residue-long hydrophobic stretches, alternating with stretches of more hydrophilic residues, can be normally inserted, provided each hydrophobic stretch is independently stable in the membrane.

In a preferred embodiment of the present invention, the extra- or intracellular domain of the membrane polypeptide contains a highly variable region, which is, upon binding to a target compound, identified and isolated as the polypeptide of interest or a section thereof. With the gene technology methods described above, in combination with combinatorial genetics, the skilled man in the art can make a library of cells displaying a huge variety of variable regions, which can be identified and isolated with the present matter in a very short time. It is possible to provide the gene for the polypeptide of interest simultaneously with the polypeptide of interest. The highly variable domain may comprise polypeptides encoded by a cDNA library or sections thereof.

The membrane polypeptide can be a homologous protein for the cell (adapted for the present method) or - which is preferred - a heterologous polypeptide for the host cells.

- 19 -

Preferably, the membrane polypeptide is genetically engineered to contain the variable domain and the selection site residing adjacent to each other on the cellular surface.

The target compound can be any material, e.g. chemically produced substances, proteins, nucleic acids, complex products derived from biological sources or mixtures or (combinatorial) libraries of these materials. Preferred substance classes for the target compound are known drugs (or derivatives thereof) (e.g. to investigate further binding patterns of such compounds to polypeptides), polypeptides (to identify and isolate protein/protein interactions), antibodies or antigens, vaccines or pathogens, etc..

If the present invention is performed using an intracellular domain of the membrane protein, i.e. the polypeptide of interest is displayed at the inner surface (including the periplasma), it is essential for the target compound to be present inside the cell. This can be achieved by transportation or diffusion of the target compound from outside into the cell (which is, however, only possible with a few or small target compounds. An alternative way of bringing also the target compound into the cell is by coexpression of the polypeptide of interest and the target compound in the same cell. Systems to enable such a coexpression of two (or more) exogenous proteins in one and the same cell are known in the art. It is clear that this way of performing the invention is restricted to target compounds that can be produced by the host cell, e.g. proteins or nucleic acids.

For selection of the interaction between target compound and polypeptide of interest, any suitable selection system known in the art may be adapted to the method according to the present invention. These selection systems are known to the skilled man in the art in principle, the adaption of any suitable selection system to the system according to the present invention will be explained below.

A critical point for the present invention is how the binding of the target compound with the polypeptide of interest influences the selection site: According to the present invention, the binding between the polypeptide of interest-portion of the membrane polypeptide and the target

- 20 -

compound results in a signal which influences the status of the selection site. This influence may be a direct or an indirect interaction of the membrane polypeptide or the bound target compound with the selection site. A direct interaction may e.g. be performed by a sterical hindrance of the selection site upon binding of the target molecule, i.e. the site, where the selection agent should bind, is not accessible anymore, because of the physical presence of the target molecule. Alternatively, such a sterical hindrance may be induced by a conformational change of the membrane polypeptide upon binding of the target compound to the polypeptide of interest-portion of the membrane polypeptide. Preferably, the selection site is present in the vicinity of the membrane polypeptide, e.g. by covalent bond to the membrane polypeptide. The selection site can also be a part of the membrane polypeptide itself. In such a case, the sterical properties of the system are clearly defined and easily adjustable by common protein engineering.

Therefore, a preferred embodiment of the present invention is a method, wherein the interaction between the extra- or intracellular domain with the polypeptide of interest and the selection site is a sterical interaction, especially a sterical interaction which does not allow the selection agent to interact with the selection site after binding of the target compound to the polypeptide of interest. Preferably, the selection site is immediately adjacent (especially via covalent bonds) to the extra- or intracellular domain of the polypeptide of interest.

In a preferred embodiment of the present invention, the selection site is an adhesion domain for the selection agent, e.g. an adhesion domain which is also part of the membrane polypeptide displaying the polypeptide of interest.

In another preferred embodiment, the membrane protein exhibits an additional domain on the surface of the cell, which additional domain is adjacent to both, the variable domain and the selection site and this additional domain also interacts, preferably by binding, with the target compound.

Indirect interaction can be performed e.g. by messenger molecules, which are released upon binding and induces a change of the selection site by binding or by contact.

- 21 -

It is essential that the change in selection specificity of the cell is in direct correlation to the binding of the target compound to the polypeptide of interest, i.e. the binding immediately induces the (direct or indirect) interaction with the selection site.

The mode of selection in the method according to the present invention is not critical, positive and/or negative selection protocols are applicable. In a preferred embodiment, the selection agent is a toxic agent.

Many of the polypeptides of the (outer) membrane are ports of adhesion for an array of toxic agents that kill or retard multiplication of the cells they enter. For the purpose of the present invention, the term "toxic agent" denotes any agent that impairs or inhibits growth of the cells. Most of the classical selection systems applied in molecular biology fall under this type of selection system. Due to its nature, this selection system is preferably applied in the „extracellular" embodiment of the present invention, wherein the polypeptide of interest is located in an extracellular domain of the membrane polypeptide.

Many toxic agents recognize a specific region (the adhesion domain) of a specific polypeptide (the receptor) on the surface of a specific cell type (the host). Binding by the agent to the adhesion domain is the first step in a process that culminates in cell lysis or inhibition of cell growth and multiplication. If this first step is impeded by mutations in the adhesion domain, by repression of the receptor, or by the interference of an inhibitor occluding the adhesion domain, a selection agent's invasion of the cell is blocked. When all of an agent's possible receptor molecules on a cell are, for whatever reason, inaccessible or dysfunctional for agent entry, the cell is rendered immune. When some but not all receptors are inaccessible or dysfunctional, the cell may become resistant but not immune: reduction of the number of receptors may slow the entry of agents, and thus reduce the rapidity or intensity of the agent's harmful effects.

For the purpose of the present invention, a selection agent may be employed that has been altered such that it is different from the corresponding naturally occurring agent with

- 22 -

regard to specificity and mode of action on the cell. For example, pathogenic agents like bacteriocins and bacteriophages - for which the above mentioned pathway via an adhesion domain is quite common - can be altered such that they change their preference for a specific adhesion domain. Thus it is easily possible for the man skilled in the art to design strategies to make pathogens adhere to a great variety of adhesion domains.

In a preferred embodiment of the present invention, the selection agent is a pathogenic agent, such as a bacteriocin, a bacteriophage or a toxin or virus of an eukaryotic cell.

Bacteriocins are bacterially produced proteins that are harmless to the strains that express and secrete them, but fatal to other, often closely related, bacteria (bacteriocins are reviewed in Pattus et al., 1990). The most extensively studied of the bacteriocins are the colicins, which are expressed by *E. coli*, and are encoded by plasmids that also code for the proteins that confer colicin immunity upon colicin-producing cells. The colicins are composed of three structural domains, each of which is responsible for one of three consecutive steps of colicin activity: i. recognition of an adhesion domain; ii. translocation across the cellular membrane; and iii. cell destruction (Pattus et al., 1990).

The colicins kill by two principal mechanisms: i. the formation of a pore in the cytoplasmic membrane, resulting in membrane permeabilization and de-energization (colicins E1, A, B, Ia, Ib, and K); ii. enzymatic cleavage of DNA or 16S ribosomal RNA (colicins E2 and E3) (Pattus et al., 1990). The kinetics of cell death when bacteria are exposed to colicins suggest that the toxic molecules enter the cell in discrete, fatal doses, which may comprise only a single molecule, but may also consist of a number of active molecules imported in one concerted step (Pattus et al., 1990). The important point here is that a single dose kills a single cell.

Almost every major outer membrane protein of *E. coli* is the receptor for at least one colicin. When multiple colicins recognize a single receptor, their adhesion domains may overlap partially. Two colicins that adhere to OmpF, a major outer membrane porin, furnish an example. As is illustrated in

- 23 -

figure 2, OmpF exhibits seven extra-cellular loops, L1 through L7, where L1 is amino proximal and L7 is carboxy proximal. The adhesion domain of colicin N includes L1, L3, L4, L5, and L6, but not L2 or L7. Thus, alterations in the peptide sequence of L1 result in total immunity to colicin N, whereas alterations to L2 do not confer even slight resistance. The adhesion domain of colicin A, another colicin that recognizes ompF, is somewhat different. Loops 1 through 5 are required for colicinA to bind its receptor, while L6 and L7 are superfluous (Fourel et al., 1990). The chimeric ompF-ompC protein 3601, which was described above, comprises loops 1-5 from ompF and loops 6 and 7 from the distinct ompC. As would be expected, *E. coli* SM1005 expressing 3601 are totally susceptible to colicin A, but totally immune to colicin N.

The number of receptors per cell is important for an easy handling of this invention's selection of specific cells from a highly variable population. The functional receptors of colicins are generally present on the surface of the cell in thousands of copies. However, it is important for an optimal performance of this invention that the level of expression of most receptors is easily manipulated: the gene encoding the receptor has already been, or may easily be, cloned into a plasmid with an inducible heterologous promoter, permitting external regulation of expression. Protocols for the construction and manipulation of plasmids with inducible promoters, permitting regulation of expression, are described in the literature (Brown, 1992, Freudl, 1989). Protocols for the construction and manipulation of plasmids with inducible promoters, permitting regulation of expression are available to those practiced in the art.

Bacteriophages are a ubiquitous and extraordinarily diverse group of viral parasites of bacteria. There are twelve families of phage, comprising 3500 distinct isolates of known morphology, and infecting over 100 bacterial genera. Their vast variation is evident even at the level of nucleic acid: depending on species, phage contain single stranded DNA, double stranded DNA, single stranded RNA, or double stranded RNA. A wide range of morphologies is usually divided into four categories: tailed; cubic; filamentous; and pleomorphic.

- 24 -

Extensive reviews of morphology, genetics, ecology, and life-cycle are available in the literature (Calendar, 1988).

Very much like bacteriocins, phages bind to specific adhesion domains of receptors on the host surface. A single outer membrane protein may be the receptor for multiple phage species, whose adhesion domains overlap to varying extents. Figure 3 shows the adhesion domains of various phages, as well as one of the colicins, that utilize *E. coli* outer membrane protein A as their receptor. The figure also shows the receptor domains that are manifestly not utilized by the phages: alterations to these regions of the receptor do not affect phage adsorption, and thus do not confer phage-resistance to the bacterial cell. Phage receptors, like those of bacteriocins, are often present in 100s to 1000s of copies per cell, but this number can be manipulated through use of a plasmid-encoded receptor regulated by an inducible heterologous promoter. As was stressed above, regulation of the level of expression of the imprinted polypeptide is important to this invention. Protocols for the construction and manipulation of plasmids with inducible promoters are described in the literature.

Following binding to the receptor, the viral nucleic acid enters the bacterial host, and phage multiplication commences. The strategies for multiplication, and the effects of multiplication upon the host, vary widely across phage species. Immediately following infection, some but not all species of phage confer immunity to subsequent infections by other individuals of the same phage species. If the infecting phage is a virulent one, it quickly modifies normal bacterial syntheses in order to appropriate them to its own multiplication and release into the extracellular space. Distinct strategies for release have very different effects upon the host. The *Inoviridae* (a family of filamentous phages) are liberated by extrusion: new phage are secreted through the bacterial membrane (Calendar, 1988). While this process is not lethal to the bacterium, it is metabolically taxing, and therefore infected bacteria divide at a slower rate than uninfected bacteria. Transmission also takes place vertically, as daughter cells inherit the infection from their parents. A

- 25 -

strategy for release similar to extrusion is exhibited by the *Plasmaviridae* (a family of pleomorphic phages). These phages, which consist only of a double stranded DNA molecule surrounded by a lipoprotein envelope, are liberated from the host by budding from the cellular membrane. Again, the process is not lethal but retards multiplication. By contrast, lysis, a means of release for tailed and cubic phages, destroys the host. After intracellular phage multiply to numbers of 20-1000, they weaken the bacterial cell from within, causing it to burst and liberate the infective particles (Calendar, 1988).

The bacteriophages exert such a wide range of effects upon their hosts that even among phage species using the same receptor there may be several different consequences of infection. Moreover, the effects of a phage upon its host can be manipulated. Mutagenesis followed by artificial selection for phage virulence or benevolence can intensify or mitigate a phage's deleterious effect on its host. Insertion or deletion of phage genes, including insertion of genes that confer metabolic capabilities or antibiotic resistance upon infected bacteria, are commonly practiced means of manipulating the impact that phage have upon their bacterial hosts. Another way to modify a phage's effects upon its host is to alter the host. For example, certain *groEL* mutations in *E. coli* produce bacterial strains that are susceptible to T4, T5, lambda, and 186 phage infection, but do not permit the intracellular assembly of mature phage particles (Zeilestra-Ryalls et al., 1993). Thus, infected bacteria are retarded, or even killed, but they do not yield infectious phage particles; in short, these bacteria are a dead-end for the phage that infect them. Finally, the effects of phage upon their hosts can be modified by the preparation of phage ghosts (phage depleted of DNA), by methods described in the literature (Boulanger et al., 1988). Because they do not carry a genome, phage ghosts do not infect cells. They do, however, adhere to cells, pierce the cytoplasmic membrane, and form ion channels, which would normally serve to transport the phage genomes. Whereas normal phage close these channels behind them, phage ghosts leave the channels open indefinitely. The rate ion efflux through these

- 26 -

channels in linearly related to the number of ghosts that bind to the cell; gradually, the efflux depolarizes the cell, which eventually dies .

Toxins and viruses of eukaryotic cells recognize specific adhesion domains of their receptors on eukaryotic cell-surfaces. The receptors of at least thirty different species of virus of eukaryotic cell lines, representing eleven different viral families, are described in the literature, and for most of these pathogens, the viral adhesion domains, which may comprise oligosaccharides, glycosylated peptide sequences, or sequences unadorned by sugars, have been characterized (Tyler et al., 1996). The receptors of eukaryotic pathogenic agents, like those of bacteriocins and bacteriophage, are commonly cloned onto vectors and expressed from inducible promoters, allowing manipulation of the number of receptors expressed per cell (Staunton et al., 1990).

Major group human rhinoviruses (HRVs) provide an example of the pathogenic agents that affect eukaryotic cells and are suitable for use in this invention. HRVs are members of *Picornaviridae*, small protein-encapsidated viruses with RNA genomes. Of 100 serotypes of rhinovirus, 89 use an extremely well-characterized glycoprotein, ICAM-1 (CD-54) as their receptor. HRV serotype 14 is well studied , easily propagated in the lab, and therefore suitable for use in this invention (Staunton et al., 1990, Couch, 1996). Major group HRVs grow efficiently in many different lines of human cell culture. The most commonly used tissues are WI-38 diploid fibroblast, fetal tonsil, MRC, HeLa and COS cell lines. Optimal conditions and protocols for growth are described in the literature (Couch, 1996). HRVs form plaques on monolayers of a variety of cell lines exposed by any one of the described techniques for exposure to viral agents. Infected cells yield 10 to 200 plaque forming units per cell after 10 to 12 hours of incubation (Couch, 1996).

The structure of HRV14 has been solved by X-ray crystallography (Staunton et al., 1990). The capsid is 30 nm in diameter, exhibiting icosahedral symmetry and a canyon, 12-30 Å wide and 25 Å deep, that runs along the 5-fold axis and contains the pathogen's receptor binding site. The receptor is

- 27 -

the well-characterized member of the immunoglobulin superfamily, ICAM-1 (CD-54), a 532 residue, 19nm x 2-3 nm rod comprising 5 immunoglobulin-like domains arranged in a line with a hinge between domains 2 and 3 (figure 4). Viral binding involves insertion of the distal immunoglobulin domain (D1) into the viral canyon. This mode of interaction is extremely vulnerable to disturbance by steric inhibition (Staunton et al., 1990). Thus, although the lower three immunoglobulin-like domains (D3-D5) are not directly involved in viral binding, and mutations within them do not disturb adhesion, their complete deletion does inhibit adhesion by reducing accessibility of D1.

ICAM-1 bears multiple N-linked oligosaccharides, and variation in glycosylation between cell-types results in ICAM-1's molecular mass heterogeneity of 76 to 114 kD. It functions as the cell-surface ligand for lymphocyte function-associated antigen 1 (LFA-1); and binding between ICAM-1 and LFA-1 contributes to leukocyte adhesion in several immunological and inflammatory processes (Staunton et al., 1990). Cytokines secreted by T-lymphocytes and monocytes, including IFN-g, TNF-a and b, and IL-a and b, induce expression of 3.5×10^6 copies per endothelial cell. However, uninduced cells exhibit much lower levels of ICAM-1, and in the CDM8 expression system, ICAM-1 can be exogenously regulated through control of an inducible promoter (Staunton et al., 1990) which makes this system excellently suitable for the function of this invention.

In a preferred embodiment of the present invention, the method is carried out in the form of a high-throughput screening assay wherein said cell population is exposed to a multiplicity of target compounds from which an unknown binding partner for a polypeptide of interest is to be identified and isolated. The target compound in this embodiment may be a single definite compound, a mixture of compounds or, to have a large variety of binding possibilities tested, a (combinatorial) library of compounds.

Such a high-throughput assay is especially suited for screening of organisms for which comprehensive or complete

- 28 -

genomic information is available, such as Mycobacterium tuberculosis or Helicobacter.

The method according to the present invention is usually carried out by performing steps a) to f) as described above. It may be necessary to repeat the selection process, e.g., with an increase in the selection pressure. This is suitable if, e.g., the number of isolated, positive clones is too large to handle.

In such cases, it is preferred to repeat steps b) through e) at least once, preferably comprising an increase in the selection pressure in the repeated steps.

The method can be carried out by steps b) through e) until the cell population comprises a desired number of distinct clones.

The term "background" is used to refer to cells that fail to bind the target compound but nonetheless remain in the selected population. In general, it is desirable to minimize the number of background cells, so that CISTEM reliably eliminates cells whose imprinted domains fail to bind the target compound, while selecting those cells whose imprinted domains do bind the target. In one strategy for the reduction of background, the CISTEM cell population is supplemented with "helper cells", whose function is propagation and amplification of the selective agent. This amplification of the selective agent increases the intensity of selection, and thereby reduces residual background cells. The helper cells lack a selectable marker that is present in the rest of the CISTEM cell population; this permits efficient elimination of the helper cells after they have served their purpose of amplifying the selective agent.

Supplemental cells may also be used in other ways to regulate the action of the selective agent. For example, to stop the action of the agent, heat-killed cells bearing the agent's receptor may be added to the CISTEM cell population. These cells bind the agent, removing them from selective activity in the remaining CISTEM cell population.

When repeating the steps, new sequence variations can be introduced in the variable regions of clones selected in step d) before steps b) to e) are repeated.

- 29 -

The screening test may also be (additionally) performed by functional parameters. In a specific embodiment of the present invention, the polypeptide of interest binds to the target compound and affects its function. Then, the polypeptides exhibited by such clones selected in step d) and amplified in step e) are screened by their ability to affect the function of the target compound.

An additional step of screening is possible, if the polypeptide of interest exhibits an enzymatic activity. In such a case, the method according to the present invention comprises a step wherein the cell population is placed in a nutrient medium in which each cell's supply of a vital compound depends upon that cell's variable region acting enzymatically upon a compound structurally similar, but not identical to the target compound.

This invention provides a method for identifying and isolating a polypeptide of interest or a target compound or a gene of said polypeptide or target using the detection of the interaction between a polypeptide of interest and a target compound. In this interaction, at least one of the interactants is unknown. The method allows for identification of the unknown interaction partner, i.e. a polypeptide of interest and/or the target compound. For some embodiments, both interactants may be unknown, e.g. when an unknown ligand, upon binding at a receptor site, exerts a certain effect on the cell which needs to be antagonized or interfered with and both the ligand and the receptor domain to which it binds need to be identified. (The words "identify" and "identification" are used here, as in the rest of the specification, to denote both the discovery of polypeptides of interest that exist in nature, and the design of polypeptides that have no natural occurrence.) Embodiments of this method are called CISTEMS, an acronym for Cellular Imprinting by Selection: Teleological Evolution of Molecules. A CISTEM applies selective pressure to a population of genetically engineered cells in order to amplify the frequency of cells in which a membrane polypeptide shows prescribed properties. This process is - as stated above - termed "imprinting" the polypeptide. The range of polypeptides that may be imprinted by this invention is broad,

- 30 -

comprising proteins, glycoproteins, proteoglycans, etc., presented in a wide range of sizes, on a diverse array of underlying structures, and by any one of several prokaryotic or eukaryotic cell cultures (figure 1).

The terms „interaction“ and „interact“ for the purpose of the present invention denote any action of the target compound upon the polypeptide of interest (in the extra- or intracellular domain of the membrane protein) and that alters or occludes the selection sitesuch that the selection agent is unable to utilize said selection site.

In the specific embodiment wherein the selection site is located on an adhesion domain, the binding of the target compound prevents the action of the selection agent via intramolecular interaction (in the membrane polypeptide), e.g. by occluding the selection agent from its adhesion domain. This may comprise binding, adhesion, structural changes or chemical modifications of the variable domain which has an impact on the adhesion domain. Such impact on the adhesion domain may, inter alia, be a conformational change of the adhesion domain that prevents the harmful action of the selection agent.

The invention is useful in at least three general areas. First, the invention may be employed to identify polypeptides with properties that are useful in medical diagnostics, therapeutics, research, laboratory procedures, or other applications. Among the polypeptides that may be identified by the method disclosed here are polypeptides that interact with, in particular bind with high affinity and specificity to, virtually any given target compound, polypeptides that adhere to their respective target compounds at multiple epitopes, polypeptides that bind their respective target compounds but show no affinity for compounds closely related to the targets, enzymatic polypeptides, and polypeptides that show biological activities, such as antagonism or agonism of a receptor. Second, by identifying polypeptide sequences that interact specifically with a target that is a known compound of significant interest, this invention may assist in the identification of the known compound's native interactants. Third, this invention may be used to analyse the interaction

- 31 -

between a known polypeptide of interest and a known target molecule. By identifying which sequences in the polypeptide of interest interact with the target molecule, the invention assists in mapping the polypeptide epitopes that play a role in the interaction. In a further embodiment, the method of the invention may be used to screen a population of target compounds and to identify those compounds that bind to a known polypeptide. In this embodiment, cells displaying a given section in the extra- or intracellular domain of the membrane polypeptide are used in a screening assay as test cells to screen for e.g. pharmaceutically active compounds that interact with the polypeptide of interest.

This invention may also be used to identify polypeptides with more sophisticated properties of interaction with the target compound than prescribed binding affinity. Among the molecules that may be identified by this method are the following: polypeptides that exhibit multivalent binding; polypeptides that bind tightly to one molecule but show no affinity for a second, closely related molecule; polypeptides that affect the functional activity of the target compound to which they bind; and enzymatic polypeptides.

The invention is based on the interaction of three CISTEM components: (1) a genetically engineered population of prokaryotic or eukaryotic cells; (2) a target compound; and (3) a selection agent (figure 5). On the surface of each cell in the population, there are - at least - two polypeptide domains involved in the operation of the CISTEM: the cellular domain carrying either a given polypeptide whose interaction with an unknown target compound is to be determined or, in the case of an unknown polypeptide, the imprinted domain, and the selection site, e.g. the adhesion domain. The imprinted domain is highly variable, because the CISTEM cell population is engineered to display up to at least 10^{12} distinct oligopeptides at the imprinted domain. (If the population is eukaryotic, the oligopeptides may be adorned by diverse oligosaccharides or other structures resulting from post-translational processing of the polypeptide). In the case of direct interaction of the target compound with the selection site, the imprinted domain on each cell's surface is

- 32 -

neighboring the selection site, e.g. on an adhesion domain, the structure to which the CISTEM's selection agent must bind in order to exert its effects upon the cell. In the case of a pathogenic agent, this may be an organism or molecule, and it may be fatal to cells or merely retardative of their growth, but the first step in the agent's destruction or retardation of a cell is necessarily binding between the agent and its adhesion domain. The two important domains - namely, the imprinted domain and the adhesion domain - are expressed in multiple copies on the surface of each cell, but they are - in the case of direct interaction - normally adjacent to each other: more specifically, they must be sufficiently close to each other that the target compound, when interacting with the imprinted domain, either extends far enough over the adhesion domain that adhesion of the selection agent is at least partially impeded, or exerts its effect on the selection site in a different way (e.g. by intramolecular steric alteration of the selection site).

The target compound is the molecule or set of molecules with which the polypeptide sequences identified by the CISTEM must interact in the prescribed manner in order to exhibit the properties or perform the functions desired of them. In other words, the CISTEM identifies polypeptides that interact with, e.g. bind to, and optionally also structurally or functionally alter, the target compound. A target compound can be any molecule, especially a protein, carbohydrate, saccharide, glycoprotein, proteoglycan, hormone, receptor, antigen, antibody, pathogen, substrate, metabolite, transition state analog, cofactor, inhibitor, drug, dye, nutrient, small molecular compound, etc., without limitation. Each CISTEM places only one requirement on the target compound: it must be effective in affecting the change in the selection site upon binding to the polypeptide of interest, e.g. that it is sufficiently large that its attachment to the imprinted domain results in occlusion of the adjacent adhesion domain or that its binding affects the structure of the membrane protein such that the selection site is sufficiently changed to prevent or to allow action of the selective agent.

- 33 -

To assemble the most basic CISTEM (exemplified for the embodiment wherein the unknown interactant is the polypeptide) and initiate selection, the following steps are performed. First, CISTEM component (1), the cell population, is incubated with CISTEM component (2), a solution or suspension of the target compound. During exposure of the CISTEM cell population to the target compound, the structural diversity of the imprinted polypeptides' variable regions results in a wide range of interactions between imprinted domains and target molecules. That is, the distinct structures displayed at the imprinted domains of separate cells exhibit characteristic on-rates, off-rates, and dissociation constants in their binding reactions with the target (figure 6). After or during incubation, CISTEM component (3) is introduced: a population of bacteriocin, bacteriophage, virus, or toxin particles is added to the cells and target compound. With this step, assembly of the basic CISTEM is completed, and selection commences.

Interaction among the three CISTEM components confers a selective advantage upon cells whose imprinted domains bind the target compound (figure 7). This takes place as follows. As described above, if the imprinted domain (in the following, when describing structural features of the invention, the term "imprinted domain" is also used for the domain carrying a known polypeptide of interest whose interaction with an unknown target compound is to be determined. Hence, the term "imprinted polypeptide" may also denote proteins carrying such known polypeptide of interest) and the selection site are immediate neighbors on the cell surface, e.g. if the selection site is on the adhesion domain, the interaction upon binding may be performed in a very definite and easily assembled manner. Consequently, when the target compound binds to a structure located at the imprinted domain, the compound can overlap the surface of the adjacent adhesion domain with the selection site. This overlap interferes with the selection agent's attachment to the adhesion domain, thus protecting the cell, and allowing it to continue dividing (figure 7). By contrast, cells whose imprinted domains do not bind the target compound with adequate affinity remain unprotected, and are

- 34 -

therefore selected, e.g. killed or retarded, by the selection agent (figure 7). Thus, each selected clone increases in frequency and in number as it continues to divide, while unprotected clones are afflicted by the selection agent. From the cell population displaying a great diversity of structures at the imprinted domain, the CISTEM selects and amplifies clones expressing structures that bind the target compound (figure 8).

Modifications and elaborations of the basic CISTEM permit selection for a prescribed affinity or nature of interaction of the imprinted domain with the target. Multivalent binding may be selected, as may highly specific binding, binding at predetermined sites, or binding by specified structural motifs. Moreover, CISTEMs may be extended to select not only for binding, but also for functional activity; these extended CISTEMs may be used to create enzymes and other functionally active molecules.

As stated above, the embodiments of the present invention are called CISTEMs, for Cellular Imprinting by Selection: Teleological Evolution of Molecules. In general, a CISTEM consists of three components. They are: (1) an engineered population of cells; (2) a target compound; and (3) a selection agent (figure 5).

The CISTEM cell population may be prokaryotic or eukaryotic. In case the unknown interactant is the polypeptide of interest, in the CISTEM cell population, one extra- or intracellular domain (the "imprinted domain") of a membrane polypeptide (the "imprinted polypeptide") is engineered to exhibit highly variable peptide sequences. If the CISTEM cell population is eukaryotic, glycosylation sites within the imprinted domain may be adorned by variable complex saccharides. High sequence variability at the imprinted domain is achieved through the insertion of variable oligonucleotides at the sites in the polypeptide gene that correspond to translated regions in which variation is desired. These variable oligonucleotides may be cDNAs, random sequences, or any other form of oligonucleotides. Techniques for inserting a gene for an exogenous oligopeptide at a targeted site in an outer membrane protein gene were described above, and an

- 35 -

extension of these techniques, which allows the insertion of variable oligonucleotides (at least 10^5 to 10^{12} different sequences, instead of a single, stereotyped nucleotide sequence), will be described in detail below.

As was described above, many pathogenic agents can kill or retard the growth of cells only if they can first bind to a specific domain on the cell surface. A CISTEM's selection agent may bind to an adhesion domain that is located immediately adjacent to the imprinted domain on the surface of the cells in the CISTEM population (figure 5). The adjacent extra- or intracellular domains in most of the membrane proteins used in this invention are separated by 3 - 30 Å. In the preferred embodiments of this invention, which are described in greater detail below, this arrangement of domains is the result of carefully choosing an imprinted polypeptide and a selection agent such that the imprinted domain is immediately adjacent to the adhesion domain. As was described above, almost every membrane polypeptide serves as the receptor for multiple selection agents, each of which utilizes a distinctive adhesion domain. However, in other embodiments of this invention, membrane proteins are genetically altered in order to construct a hybrid polypeptide in which the imprinted domain is placed beside the adhesion domain. The techniques for engineering membrane proteins, including the fusion of two separate polypeptides to form a chimeric polytopic polypeptide, were described above; their specific applications to embodiments of this invention are described below.

The target compound is any molecule or set of multiple, distinct molecules, which are simply added to the cell population according to the present invention or, in some embodiments of this invention, are secreted by the cells themselves. The present invention identifies and selects cells in which the displayed polypeptide sequence interacts with the target compound, or some subset of molecules included in the target compound, and this interaction event is necessary to the functions to which the polypeptides are ultimately applied, such as enzymatic activity or alteration of the target compound's biological activity. In the application of

- 36 -

this invention to the analysis of molecular interactions, the target compound may be a known compound of interest that is thought to have important interactions with unknown polypeptides. Identification of these unknown polypeptides is assisted by this invention's identification of sequences with which the target compound interacts. In epitope mapping, the target compound is a molecule of interest (not necessarily a polypeptide), which is known to be involved in an interaction with a polypeptide of interest. By identifying polypeptide sequences that bind the target, this invention assists in delineating the regions of the known polypeptide that are directly involved in its interaction with the target compound. For each of the three areas of applications that have been mentioned here - identification of sequences with useful functions; interaction analysis; and epitope mapping - specific target compounds are described hereinafter.

The third component of the invention is the selection agent. The most important requirement of the selection agent has already been stated: the agent's effects upon cells must be contingent upon the agent's ability to bind to a selection site which is modified upon the binding event of the target compound. This selection site is preferably an adhesion domain that is adjacent to the imprinted domain. Other attributes of the selection agent, e.g. an pathogenic agent, such as virulence and concentration, may be altered, if necessary, by common laboratory procedures. These attributes and their adjustment have already been described.

There are two steps in the execution of selection for polypeptide sequences that bind the target compound. In most embodiments of this invention, these steps are performed sequentially, though they may in some cases be simultaneous. First, the CISTEM cell population, component (1) of the invention, is incubated with a suspension or solution of the target compound, component (2) of the invention. The solution of incubation is a standard medium for cell growth, described in more detail in the passages relating to cell lines and media. During incubation, the structural diversity of the variable regions results in a wide range of binding interactions between imprinted domains and target molecules:

- 37 -

the distinct structures displayed at the imprinted domains of separate cells exhibit characteristic on-rates, off-rates, and dissociation constants in their binding reactions with the target (figure 6). Incubation continues until the binding between imprinted domains and target compounds approaches equilibrium levels. For most embodiments of the invention, one hour is sufficient. In certain embodiments of the invention, the period of incubation is cut short in order to favor polypeptides that exhibit not only a high affinity for the target, but also exceptionally high on-rates; these CISTEMS are described below.

After incubation, the selection agent, component (3) of the invention, is added to the suspension of cells and target compound. Exposure of the cell population is performed by a method appropriate to the agent. For example, if the agent is a bacteriophage, the suspension of cells and target compound is exposed to classical phage infection. Alternatively, if the agent is a colicin, exposure of the cell population to the agent is performed simply by adding the colicin to the medium in which the cells are incubated with the target compound. When the selection agent in the form of a pathogenic agent is a reproducing organism, such as a bacteriophage, it may be necessary to stop its action after a period of exposure. The appropriate concentration of pathogenic agent, period of exposure, and procedures for halting the action of the agent are in the general knowledge of the skilled man in the art.

Addition of the selection agent to the suspension of cells and target compound completes assembly of the CISTEM, and selection commences. Interaction among the three CISTEM components confers a selective advantage upon cells whose imprinted domains bind the target compound: e.g. when a structure located at the imprinted domain binds the target compound, the compound overlaps the surface of the adjacent adhesion domain. Because the distance between adjacent extra- or intracellular domains of the polypeptides used in preferred embodiments of this invention is only 3 - 30 Å even a small peptide hormone is a sufficiently large target to reach from the imprinted domain, where it is bound, to overlap the abutting adhesion domain. This overlap interferes with the

- 38 -

selection agent's attachment, protecting the cell and allowing it to multiply (figure 7). Cells whose imprinted domains do not bind the target compound with adequate affinity remain unprotected, and, consequently, they are killed or retarded by the pathogenic (selection) agent (figure 7).

Each cell in the CISTEM population expresses multiple copies of its imprinted polypeptide, and binding is a dynamic process balancing on-rate and off-rate. Therefore even when a cell's imprinted polypeptide exhibits high affinity for the target compound, some copies of the imprinted polypeptide are left unprotected some of the time. Consequently, every cell in a CISTEM population, no matter how high its affinity for the target compound, has some probability of being hit by a selection agent. Nonetheless, this probability is lower when a cell's affinity for the target compound is higher.

Furthermore, because each clone in a CISTEM is present in multiple copies (the CISTEM cell population is redundant) the probability that a single cell will be hit is also the fraction of that cell's clonal population that will be hit, and the differential elimination of clonal populations is the very basis of selection. Thus, each selected clone increases in frequency and in number as it continues to divide, while unprotected clones are destroyed. In short, the CISTEM selects and amplifies clones expressing structures that bind the target compound (figure 8).

The effect of the target compound relevant for the preferred embodiment of the invention, wherein the selection site is located on an adhesion domain is that, due to its size, it extends, upon interaction with the imprinted domain, from the imprinted domain to the adhesion domain. This is 3 - 30 Å for many imprinted polypeptides imprinted by this invention. Alternatively, independent of its size, the interaction between the target compound and the imprinted domain may lead to a conformational and/or chemical change of the variable domain and/or the adhesion domain which prevents the selection agent from attachment. The target compounds, their source, preparation, and use of CISTEMs with particular applications in this specific embodiment of the invention are described hereinafter. While the concentration of target

- 39 -

compound in a CISTEM is also important to its function, this concentration depends on the nature of other CISTEM components, and is therefore discussed below.

In CISTEMS designed to imprint polypeptides to serve as highly specific markers or partitioning agents in diagnostic assays, the target compound is simply the molecule that is to be tagged and/or partitioned by the polypeptide. For example, in an embodiment of the invention that is used in the design of a diagnostic polypeptide that binds tightly and specifically to Human Immunodeficiency Virus (HIV), the target compound is HIV gp120, the extracellular subunit of a major HIV glycoprotein (Fantini et al., 1997). The production of recombinant gp120 in a soluble form that is suitable for use in CISTEMS is described in the literature (Fantini et al., 1997). Another embodiment of the invention is used to design a diagnostic polypeptide for the direct detection of a different virus, Human T-cell Leukemia Virus type I (HTLV-1); here the target compound is the HTLV-1 surface glycoprotein gp46. The production of authentic, glycosylated HTLV-1 gp46 in a recombinant vaccinia virus/T7 polymerase system is described in the literature (Arp et al., 1996). Diagnostic assays for detecting human viral infection often bind not to the virus itself, but rather to the antibodies raised against the infectious agent. One preferred embodiment of this invention imprints a polypeptide to exhibit high affinity for a monoclonal antibody, MAbT7. The target compound in this CISTEM, MAbT7 itself, is commercially available (Novagene).

The invention is also used to imprint polypeptides to serve as antagonists, agonists, or modulators of the functional action of biologically active molecules. When this invention is used to imprint a polypeptide to serve as an antagonist, the target compound may be either of the molecules involved in the interaction to be antagonized. HIV gp120 is the viral glycoprotein that binds to the CD4 receptor, and thus mediates viral invasion of the human cell (Fantini et al., 1997). The embodiment of this invention that imprints a polypeptide (ICAM-1, which is described above) to bind specifically to HIVgp120, provides a polypeptide that could

- 40 -

serve as an antagonist of the interaction between the viral glycoprotein and its host cell receptor.

The polypeptides selected by a CISTEM are more likely to functionally antagonize a given molecular interaction when the CISTEM's target compound is pared down, so far as is possible, to one interactant's active site. Structures on either interactant that are distinct and topologically distant from the active site are less promising as target compounds because a polypeptide that binds such structures may not interfere with the normal interaction of the active molecules. For example, one embodiment of this invention imprints a polypeptide (*E. coli* OmpA, which is depicted in figure 13, and described further below) to function as an antagonist of the interaction between the eukaryotic cell-signalling proteins Ras and Raf (Block et al., 1996). (Activation of Raf, a protein kinase, by the immediately upstream protein, Ras, initiates pathways leading to the phosphorylation of transcription factors, the stimulation of numerous genes, and, ultimately, to cell proliferation and differentiation. Constitutively active oncogenic mutants of Ras or Raf occur in a large number of human tumors (Block et al., 1996).) In this CISTEM, the target compound is the domain of Raf (amino acids 51-131) that is known to be directly involved in binding to Ras (Raf's so-called Ras-binding domain, or Raf-RBD) (Block et al., 1996). The construction of a recombinant expression vector for Raf-RBD, expression in *E. coli*, and purification of the protein are described in the literature (Block et al., 1996). Using Raf-RBD, rather than Raf in its entirety, as the target compound ensures that the polypeptides selected by the CISTEM bind to precisely that domain of Raf which interacts with Ras, rather than to Raf epitopes that are extraneous to the important cell-signalling interaction.

When this invention is used to imprint polypeptides to serve as agonists, rather than antagonists, of a receptor or biological pathway, the target compound is the receptor or biomolecule that the imprinted polypeptide is intended to activate. For example, in an embodiment of this invention that imprints a polypeptide (*E. coli* OmpA) to mimic interleukin 6 and function as an agonist of the interleukin 6 receptor (IL-

- 41 -

6R), the target compound is a solubilized form of IL-6R: the expression of soluble human IL-6R in CHO cells, and the protein's purification, are described in the literature (Yasukawa et al., 1990). After this CISTEM selects polypeptides with affinity for IL-6R, a secondary screen is used to isolate polypeptides whose binding to the receptor results in transduction of the signal that is normally initiated by IL-6. This secondary screen is described in the passages relating to extensions of the basic CISTEM.

When this invention is used to create agonists, the target compound is pared, so far as is possible, to the active site of the receptor or biomolecule that is to be agonized. This is done because removal of extraneous epitopes from the target compound increases the probability that an imprinted polypeptide that binds the target will also be active as an agonist.

In another example of this invention's application to the creation of agonists, CISTEM is used to identify, or to create mimics of, specific antigens that are present in infectious pathogens or cancer. These antigens or antigen mimics are then often used to provoke a specific immune response against the pathogen or tumor. In these embodiments of the invention, the target compound is composed of more than one molecule: multiple monoclonal antibodies, specific immune receptors, such as cytotoxic T-lymphocyte receptors or MHC-I molecules, or polyclonal antisera derived from individuals exposed to a specific antigen, such as meningococcal Opa proteins, to a specific pathogen, such as meningococci, or to a tumor, such as melanoma. Polyclonal antisera are divided from animals, human volunteers or actual patients. This composite target is used as an impression - a "negative", or "footprint", of the antigen - and from this footprint, the CISTEM is used as described in example 2 to identify the antigen or to create a mimic of the antigen. An antigen-mimic is of interest, because a polypeptide that is imprinted to bind the antibodies composing sera against an antigen is likely to elicit an immune response very similar to the one provoked by the antigen. Antibodies suitable for use as the target compound in a CISTEM for imprinting *E. coli* OmpA to mimic Neisserial

- 42 -

protein Opa 5c include mAb B306, mAb A222/5b, and others described in the literature (Olyhoek et al., 1991). Antibodies for use as the target compound in a CISTEM for identifying tumor antigens are derived from the polyclonal serum of a cancer patient, the isolation of which is widely described in the literature.

In order to create catalysts of chemical reactions, such as ester hydrolysis, polypeptides are imprinted to bind to transition state analogs. The preparation and structure of these transition state analogs are described in the literature (e.g. Charbonnier et al., 1995). In one embodiment of this invention, a phosphonate transition-state analog (structure and preparation described in Charbonnier et al., 1995) is used as the target compound for imprinting *E. coli* OmpA to function as a catalyst of the hydrolysis of p-nitrophenyl esters.

In other embodiments of this invention that are used to create polypeptides with specific enzymatic activity, the target compound is not a transition state analog, but rather a modified substrate. The target is a molecule that is structurally very similar, but not identical, to the substrate upon which the imprinted polypeptide will ultimately act. The reason for using a slightly modified structure, rather than the substrate itself, is the following: if the substrate itself were used as the target compound, imprinted polypeptides that not only bound the substrate, but also performed the desired enzymatic activity, would alter the substrate and then release it, rendering the underlying cell sensible to the selection agent. Thus, if the substrate itself were used as the target compound, the CISTEM would select against clones expressing polypeptides with the desired enzymatic activity. Therefore, the target compound must be an effective structural mimic of the substrate (i.e. polypeptides that bind the target must also be able to bind the substrate), but it must be invulnerable to the prescribed enzymatic activity (i.e. polypeptides that act enzymatically on the substrate must be unable to act on the target compound).

For example, in an embodiment of this invention that imprints *E. coli* OmpA to function as a sequence-specific protease that hydrolyzes the bond between two methionines, the

- 43 -

target compound is the oligopeptide $\text{NH}_2\text{-R1-R2-R3-M1=M2-R4-R5-COOH}$. M1 and M2 represent methionine residues; the bond between them, which is represented by "=", is a nonhydrolyzable methylene-group bond. Normal amide bonds are represented by "-". R1...5 are variable residues, which are randomly chosen from the 20 amino acids. These residues are varied so that only imprinted polypeptides that bind to the consistent M1=M2 motif of the target compound are selected by the CISTEM. Imprinted polypeptides that show high affinity for a different peptide motif bind at most only a tiny fraction of the target polypeptides, and are therefore left vulnerable to the selection agent. Several methods for the synthesis of this oligopeptide target compound are described in the literature (Janda, 1994).

After the CISTEM has selected clones expressing imprinted polypeptides that bind the modified substrate, a secondary screen may be applied to isolate imprinted polypeptides that not only bind the target, but also perform the desired enzymatic reaction. Secondary screens, including the one used in the CISTEM that imprints polypeptides to hydrolyze the bond between two methionines, are described below.

In addition to the creation of functionally active polypeptides, such as antagonists and enzymes, this invention's applications include interaction analysis (current techniques reviewed in Phizicky et al., 1995). By screening a diverse library of polypeptides for sequences that interact specifically with a target compound of significant interest, this invention assists in the identification of the compound's native interactants. For example, although the structural interaction between Ras and Raf-RBD has been well-characterized through crystallography and mutational analysis (Block et al., 1996), the activation of Raf by Ras is a complex process that is not yet completely understood, and probably involves interaction with other molecules. The CISTEM that screens polypeptides displayed in OmpA for interaction with Raf-RBD assists in the identification of polypeptides involved in Raf activation. This CISTEM involves the following components: Loop IV of E.coli OmpA is used to display the polypeptides translated from fragments of cDNAs derived from a

- 44 -

cell expressing Ras and Raf (the insertion of foreign oligonucleotides into the gene for OmpA is described below); Raf-RBD (see above for a more detailed description of this target compound and its preparation) serves as the target compound; and phage K3 serves as the (pathogenic) selection agent. The appropriate concentrations of the target compound and the selection agent can easily be determined by the skilled man in the art after reading the present specification and are described below in the examples of OmpA imprinting CISTEMS.

This invention is also used for epitope mapping (Phizicky et al., 1995), serving to elucidate the surface of structural interaction between two molecules. In CISTEMS for epitope mapping, the target compound is one of the two molecules whose interaction is to be mapped. Usually it is the interactant that is most easily prepared in soluble form that serves as the target compound. Thus, if one of the two interactants is a membrane protein, it is often the other molecule that serves as the target. In a CISTEM used to map the epitopes of interaction between Ras and Raf-RBD, Raf-RBD serves as the target compound. As noted above, it is readily prepared in soluble form (Block et al., 1996). Thus, one of the interacting proteins is dissected into multiple smaller domains which are displayed in the imprinted domain and selected by the CISTEM for their capability to bind to the other interactant.

In certain embodiments of this invention, multiple separate molecules compose the target compound. Each molecule must meet the CISTEM's size requirement and must be present in the CISTEM at a concentration sufficient to confer protection upon clones that bind to it with high affinity. (Selection by CISTEMS with multiple distinct target molecules strongly favors binding by imprinted domains to multiple distinct molecules, because the effective concentration of target compound rises significantly with each additional compound that the imprinted domain is capable of binding.) When the target compound is a library of molecules and the imprinted domain displays variable regions, the CISTEM assists in the identification of interacting pairs of molecules: polypeptides

- 45 -

expressed in the imprinted domain are selected for their affinity for particular members of the library of molecules that composes the target compound. Those cells that express an imprinted polypeptide that binds to one or more molecules in the target compound are selected, while cells expressing an imprinted polypeptide that fails to bind any component of the target compound are eliminated by the CISTEM's selection agent. In certain embodiments of this invention, which are described below, the target compound consists of a library of polypeptides or phage-displayed polypeptides that are secreted by the CISTEM cell population itself.

In an embodiment of the invention, the target compound is the interaction partner that is to be identified due to its interaction with a given polypeptide of interest, e.g. a section of a naturally occurring cell surface receptor which is displayed in the extracellular domain. In this case, the target compound is the unknown interactant. In this example, the cells presenting the receptor domain of interest may be incubated with chemical compounds from a pool of substances. Surviving cells are indicators that the compound that has interacted with these cells is a candidate substance for interaction with the receptor domain.

In a further embodiment of the invention cells are protected from the (pathogenic) selection agent in an autocrine fashion, due to binding between the imprinted domain and a polypeptide target that is secreted by the cell itself. In this embodiment, two separate libraries of polypeptides, expressed from two separate libraries of genes, are screened for interaction with each other. The cell population displays one library on the cell surface, exactly like the standard CISTEM cell population. In this embodiment, however, the cell population carries also a second library of genes. These genes reside in a vector that encodes a secreted polypeptide. Such vectors are described in the literature (Mollenkopf et al., 1996). Each cell in the autocrine population displays one foreign polypeptide on its surface and secretes a second foreign polypeptide into the extracellular medium. In the event that these two polypeptides bind to each other, a cell

- 46 -

protects itself from the pathogenic agent and is therefore positively selected.

As was noted above, the term "polypeptide" relates to all types of peptides or proteins synthesizable by a host cell. To be imprinted by this invention, a polypeptide must comprise at least one extra- or intracellular domain that accepts exogenous oligopeptide insertions without disturbing translation or export of the polypeptide to the cell surface. This capacity to accept insertions has been observed in numerous secreted polypeptides, which can be attached to the outer membrane by a genetically engineered anchor that is described in the literature (Singer, 1990; Cross, 1990), as well as in each one of the dozens of bacterial, yeast, and mammalian membrane polypeptides that have been topologically mapped by gene fusion techniques. A number of these polypeptides, such as *E. coli* OmpA, *E. coli* LamB, and mammalian ICAM-1, their properties, and their capacities to exhibit exogenous sequences in extracellular domains, are described in the literature.

The membrane polypeptide may be autologous or heterologous for the cell. Preferably, the membrane polypeptide is heterologous. In preferred embodiments of this invention, all copies of the membrane protein gene in a given clone must be altered in some way. For example, if a plasmid-borne copy of the gene for the imprinted polypeptide bears a foreign insert at the imprinted domain, whereas the chromosomal copy of the gene remains wildtype, then proper selection by the CISTEM is hindered, because the selection agent and/or the target compound can bind to the wildtype copy of the polypeptide, as well as to the engineered copy. In case the cell contains the autologous membrane polypeptide but otherwise has the desired properties for its use in the present invention, the gene encoding the membrane protein is inactivated, e.g. by deletion. Standard molecular methods are available to the person skilled in the art.

Exogenous insertions at an extra- or intracellular domain are involved in two stages of each CISTEM. First, an oligopeptide tag is inserted into the imprinted domain, and cell populations exhibiting this tag are used in preliminary

- 47 -

trials to optimize the CISTEM. The oligopeptide tag, its insertion into the imprinted domain, and construction of cell populations exhibiting the tag, are all described in this section. Second, once preliminary trials with the tagged cell population have been performed, the actual CISTEM cell population is constructed: in this population, variable peptide sequences, rather than an oligopeptide tag, are inserted into the imprinted domain. The steps in the construction of CISTEM cell populations are also described here.

By conventional methods familiar to those practiced in the art, the gene encoding the imprinted polypeptide is cloned onto an appropriate vector. The vector's level of expression is important to CISTEM function. In most CISTEM cell populations, roughly 1000 copies of the imprinted polypeptide are exhibited per cell. However, as is described below, in some cases expression is adjusted to achieve proper CISTEM selection. In preferred embodiments of this invention, regulation of the level of expression of the imprinted polypeptide is made possible through the use of a vector comprising an inducible promoter, such as the prokaryotic lac regulatory elements (Freudl, 1989), which are inducible with isopropyl-b-D-thiogalactopyranoside (IPTG), or the eukaryotic interferon promoter. The vector also contains a selectable marker, such as a gene for antibiotic resistance, to facilitate transformation (selectable markers and their use in cloning vectors are extensively described in the literature, e.g., Brown, 1992, Freudl, 1989).

For prokaryotic proteins, common expression plasmids are used, and the polypeptides imprinted in preferred embodiments of this invention have already been placed on suitable plasmids. For example, on pEV218 (Freudl, 1989), the plasmid used in a particularly preferred embodiment of this invention, *E. coli* OmpA (*E. coli* Outer membrane protein A is described) has been placed under the inducible control of the lac regulatory elements residing on the common, high-copy-number plasmid pUC9. Expression of ompA on pEV218 is further controlled by the presence of a TAG stop codon corresponding to amino acid 7 of the polypeptide: when the plasmid is used

- 48 -

to transform an amber-suppressing host, such as *E. coli* UH203 (*supF*), expression is regulated not only by induction with IPTG, but also by the concentration of the suppressor tRNA. The plasmid also comprises a gene for ampicillin resistance. Similarly, on pSB1649 (Brown, 1992), *E. coli* *lamB* is regulated by the inducible *lac* promoter on the common plasmid pAC1. For eukaryotic polypeptides, common viral vectors are suitable, and many polypeptides have been cloned into appropriate vectors. For example, for the polytopic mammalian glycoprotein (ICAM-1) (Intercellular Adhesion Molecule-1 is described further below), which is imprinted in a preferred eukaryotic embodiment of this invention, the CDM8 cDNA expression vector pCD1.8 may be used in conjunction with the COS cell expression system (Staunton et al., 1990).

The vectors bearing the gene for the imprinted polypeptide are replicated and isolated. They may be replicated by a cell population, by PCR techniques, or by any other commonly employed methods described in the literature. In preferred embodiments of the invention, 10^8 to 10^{10} copies of the vector are isolated. The exact number depends on the number of distinct clones that will compose the CISTEM cell population, and on the technique used in the step described below, the insertion of the variable oligonucleotides into the vectors.

The gene for a well-characterized oligopeptide tag is inserted into the vector at a genetic site corresponding to the imprinted domain of the polypeptide. Alternatively, the oligonucleotide tag may replace a nucleotide sequence of the naturally occurring sequence. Replacement is accomplished by excising a natural coding region and by inserting the oligonucleotide tag in place thereof. The tag may be any peptide sequence that is known to bind with high affinity and specificity to a ligand that is commercially available or easily produced in the laboratory by techniques described in the literature. One example of a commercially available tag/ligand pair is the T7 sequence (*Met-Ala-Ser-Met-Thr-Gly-Gly-Gln-Gln-Met-Gly*), which binds to the T7 antibody (available from Novagene, TB15 protocol). An example of a tag/ligand pair that is described in the literature is amino

- 49 -

acid sequence 36-41 of the eukaryotic polypeptide Ras and the Ras-binding domain (amino acid sequence 51-131) of Raf (Block et al., 1996). The production and purification of Ras-RBD by commonly practiced techniques is described in the literature (Block et al., 1996). Many other tag/ligand pairs, such as peptides that bind to specific antibodies, are also described in the literature and available to those familiar with the art (Phizicky et al., 1995). To insert the gene for the oligopeptide tag at a genetic site encoding the imprinted domain, any common technique for inserting an oligonucleotide at a specific site in a vector can be used (e.g. Boyd, 1994). Several such techniques are referenced above, and one preferred procedure is described here in more detail.

Many membrane polypeptides have been topologically mapped, and these maps are available in the literature. Using a topological map of a membrane polypeptide's sequence, two restriction endonuclease sites (site R1 and site R2) each of which is unique in the vector and is located in the genetic region coding for the imprinted domain, are identified. If two unique restriction sites do not already exist at the desired site of insertion, they may be engineered using any one of a number of procedures described in the literature (examples of unique restriction sites being engineered at a desired site of insertion are described in Brown, 1992; Freudl, 1989). The vector is cleaved at the two designated sites by digestion with restriction endonucleases R1 and R2, following the protocols and under the conditions specified by the suppliers of the enzymes. This yields a linearized vector with non-complementary sticky ends (RS1 and RS2) and an excized fragment. The linearized vector is isolated by standard electrophoretic size fractionation.

An oligonucleotide template comprising the gene for the tag is synthesized using conventional technology, such as phosphoramitide chemistry, or can be obtained from commercial laboratories offering custom oligonucleotides. The oligonucleotide is synthesized to read 5'-P--R1--tag gene--R2--3', in which P is the primer sequence for a proof-reading polymerase, such as Vent polymerase (New England Biolabs), R1 and R2 are the restriction endonuclease sites mentioned above,

- 50 -

and --tag gene-- is the genetic sequence encoding the oligopeptide tag. (It may be necessary to include one or two extra-oligonucleotides flanking the tag gene so that the tag is translated in frame with respect to the imprinted polypeptide.) Using the polymerase whose primer is P, the oligonucleotide template is used to generate a double stranded oligonucleotide. This oligonucleotide is digested with restriction endonucleases R1 and R2 under conditions described by the enzyme suppliers, yielding an oligonucleotide with sticky ends RS1' and RS2'. This oligonucleotide, which has sticky-ends complementary to the linearized vector, is ligated into the vector by a common ligation procedure familiar to those practiced in the art. The result is that the vector's gene for the imprinted polypeptide bears the gene for the oligopeptide tag at a genetic site corresponding to the imprinted domain.

An alternative procedure for the construction of the tag gene with sticky ends RS1' and RS2' is used in some preferred embodiments of this invention. In this procedure, two semi-complementary single stranded oligonucleotides are synthesized by conventional technology or acquired from a laboratory offering custom oligonucleotides. The two single stranded oligonucleotides are synthesized such that, when they are annealed, they form a double-stranded oligonucleotide with sequence RS1'--tag gene--RS2'. This double stranded oligonucleotide is then ligated into the linearized vector described above.

An example of this procedure is provided by a preferred embodiment of this invention, which involved the construction of an *E. coli* population displaying the T7 tag (figure 9) at loop IV of OmpA. In pEV218 (Freudl, 1989), the ompA gene has been engineered to bear a string of restriction endonuclease sites at a genetic region corresponding to extracellular loopIV of the polypeptide (figure 9). This plasmid was digested with SacI (enzyme and reaction buffer from New England Biolabs) and XbaI (enzyme and reaction buffer from MBI Fermentas), yielding a fragment of the restriction site polylinker and the linearized plasmid. The linearized plasmid was isolated by standard gel electrophoresis. Two single

- 51 -

stranded oligonucleotides were synthesized by conventional technology to read as follows:

Oligonucleotide 1:

5' - CG[ATGGCTAGCATGACTGGTGGACAGCAAATGGGT] - 3'

Oligonucleotide 2:

5' - CTAG[ACCCATTTGCTGTCCACCAGTCATGCTAGCCAT]CGAGCT - 3'

The sequence in brackets comprises the gene segment encoding the oligopeptide that is recognized by the T7 antibody. These two single-stranded oligonucleotides were annealed in MBI Fermentas Klenow buffer, forming the double stranded T7-tag gene segment with sticky ends complementary to the sticky ends on pEV218, linearized by SacI and XbaI, as described in the previous paragraph. When the double-stranded oligonucleotide was ligated into the linearized plasmid, the T7 tag gene segment was in frame with the ompA gene on pEV218. The new construct was designated pEV218T7.

After the tag gene has been inserted into the vector, the vector is used to transform a population of cells lacking a chromosomal copy of the gene for the imprinted polypeptide. Strains lacking chromosomal genes for outer membrane polypeptides are described in the literature (e.g., Brown, 1992, Freudl, 1989, and Staunton et al., 1990), as well as below, in the passages relating to cell lines and media. Conventional methods of transformation are used, and transformants are selected by their expression of the vector-borne selectable marker, such as a gene for ampicillin resistance. In a particularly preferred embodiment of this invention, in which *E. coli* OmpA was engineered to display the T7 polypeptide tag at loopIV, pEV218T7 was used to transform *E. coli* strain UH203 (*lac*, *supF*, *ompA*, *recA*, *proAB*, *rpsL*) (Freudl, 1989), according to a common method of electrotransformation, as follows.

25 ml of overnight culture of UH203 (Luria broth, 37°C) was suspended in 1l of LB. Cells were cultivated at 37°C, with shaking. When the culture reached OD₆₀₀ = 0.5, the cells were chilled for 15 min. on ice, and then centrifuged in a cold rotor at 4000 x g for 15 min at 4°C. Supernatant was removed. Sediment was gently resuspended in 1l of ice-cold 10% glycerol and centrifuged as before. This washing procedure was repeated

- 52 -

three times and cells were resuspended in 1 ml of ice-cold 10% glycerol. 42µl of the cell suspension was mixed with 2µl of DNA (pEV218T7 described above) solution (1pg dissolved in TE) and incubated 10 min on ice. 40µl of this suspension was pipetted to an electroporation cuvette (BIO-RAD Gene Pulser with Pulse Controller was used. Apparatus settings were as follows: 2,5 kV; 200 W; 25µF; time constant = 4,9 ms. Immediately following electroporation, cells were washed in 600µl of SOC medium and transferred to a cultivation tube. Cells were then allowed to recover for 1 hour at 37°C in a shaking flask. Finally, cells were plated on LB plates with 100µg ampicillin/ml.

There are several ways to generate the variable oligonucleotides that are inserted into the vector, and, ultimately, translated into the variable regions exhibited by the CISTEM cell population. To facilitate insertion into the vectors, the variable oligonucleotides must have conserved sequences at their termini, just as the tag gene described in the previous section was given defined termini RS1 and RS2. For example, in preferred embodiments of the invention, variable oligonucleotides with defined sequences at their termini are easily synthesized using solid-phase phosphoramidite chemistry (Brown, 1992): at defined terminal positions, only one of the four nucleotides is made available for addition to the strands being synthesized, whereas at variable positions, a mixture of multiple nucleotides is presented to the strands. Alternatively, end addition of bases, catalyzed by terminal transferase in the presence of nonlimiting concentrations of multiple nucleotides, can append a highly variable sequence to a defined terminal. Highly variable sequences may also be obtained by selecting fragments of defined length from digestions of natural cellular DNA or RNA preparations.

In a CISTEM for screening the polypeptides encoded by a cDNA library, the variable oligonucleotides are the cDNA library genes. The size of the cDNA library that is screened by such a CISTEM is in some embodiments of this invention reduced by limiting the screen only to genes of interest, as determined by comparison of the cDNA library genes to

- 53 -

sequences catalogued in genomic libraries. In a CISTEM for identifying cancer antigens, cDNAs are prepared from tumor cells according to protocols described in the literature.

In a CISTEM for analyzing the interaction between a known polypeptide and a second known compound of interest, the variable oligonucleotides are sections of the gene encoding the known polypeptide. For example, in a preferred embodiment of this invention, in which the interaction between the eukaryotic cell-signalling proteins Ras and Raf is analyzed, the oligonucleotides inserted into pEV218 are sections of the sequence encoding Ras.

In preferred embodiments of the invention, stop codons are excluded from variable oligonucleotide inserts in order to prevent truncation of the imprinted polypeptide. This is achieved by synthesizing oligonucleotides to read $D1(NNP)_nD2$, where D1 and D2 represent the defined sequences that facilitate insertion of oligonucleotides into vectors; n represents the number of triplet codons that compose the variable section of each oligonucleotide; N represents all four nucleotides, each with equal probability; and P represents a subset of nucleotides, each with equal probability. NNP encodes all 20 amino acids, but precludes the occurrence of stop codons recognized by the CISTEM cell population. Thus, when P is allowed to be G, C, or T, the stop codons TAA and TGA are excluded from the oligonucleotide inserts. It is not necessary to preclude the occurrence of all three stop codons in inserted oligonucleotides, because the CISTEM cell population may be composed of stop codon-suppressing mutants, which are capable of translating one or more of the stop codons. Numerous suppressor strains are described in the literature (e.g. Brown, 1992, Freudl, 1989).

When the imprinted polypeptide is a glycoprotein or proteoglycan, the inserted oligonucleotides may include consensus sequences for N-linked glycosylation (Chavez et al., 1991). These are synthesized as defined sequences interrupting the variable sequence of each oligonucleotide. In CISTEMs that incorporate an element of structural design, variable oligonucleotide inserts may include a mixture of defined and

- 54 -

variable positions, sequences coding for fixed motifs, or statistical biases toward certain sequences or structures.

The variable oligonucleotides are inserted at a prescribed site in the vectors constructed and amplified by the methods described above. Two preferred procedures for precisely targeted insertion are described here—one in which the number of oligonucleotides inserted into each vector is defined, and another in which the number of inserts is variable. The available procedures for inserting oligonucleotides at specific vector sites are certainly not limited to the two described here. In fact, any of the gene fusion techniques used to insert a gene coding for a reporter moiety into the gene for a membrane protein may be used to the same effect as the procedures that will be described.

The site of insertion in the vector is located such that oligonucleotides inserted there are translated as part of the imprinted domain. The site of insertion may be further specified (i.e. within the sequence encoding the imprinted domain) according to structural analysis or prescriptions for a particular nature of interaction between the imprinted domain and the target compound. To insert oligonucleotides, both of the procedures outlined here make use of restriction endonuclease sites that are unique in the vector, and are located at or near the prescribed site of insertion. If a unique restriction site does not already exist at the site of insertion, one may be engineered using any one of a number of procedures described in the literature (e.g. Boyd, 1994, Brown, 1992 and Freudl, 1989).

To insert one and only one variable oligonucleotide per vector, two distinct restriction sites, each unique in the vector, must be located such that they flank a short restriction fragment comprising the prescribed site of insertion. If the sticky ends left by cleaving the vector with one endonuclease (R1) at its unique site are RS1 and RS1', while cleaving with the other endonuclease (R2) at its unique site leaves sticky ends RS2 and RS2', then the variable oligonucleotides generated at step 3 are given defined and conserved sticky ends RS1 and RS2'. The vectors are digested by both restriction nucleases (R1 and R2), and the digestion

- 55 -

products are size-fractionated to isolate the vectors, each of which has non-complementary sticky ends RS1' and RS2. The population of restriction-digested vectors is combined with a slightly smaller population of variable oligonucleotide inserts (vectors/insert » 1.5 - 2.0), and ligation is allowed to occur. Because the sticky ends of the vectors are non-complementary to each other, each vector can neither ligate shut, nor ligate to other vectors. Similarly, because oligonucleotide inserts have non-complementary sticky ends RS1 and RS2', they can neither self-ligate into loops, nor ligate to each other to form concatamers of multiple oligonucleotides. The only ligation that can occur is the insertion of one variable oligonucleotide into each vector. After ligation has taken place, the vector population is size-fractionated to discard vectors that did not receive an insert. In the resultant vector population, each vector has one variable oligonucleotide at the prescribed site in the region encoding the imprinted domain.

An example of this procedure for inserting one and only one variable oligonucleotide per vector is furnished by a preferred embodiment of this invention, in which a cell population is engineered to display variable oligopeptides at loop IV of *E. coli* ompA. In pEV218 (Freudl, 1989), the ompA gene has been engineered to bear a polylinker rich in restriction endonuclease sites at a genetic region corresponding to extracellular loopIV of the polypeptide (figure 14). This plasmid is digested with SacI (enzyme and reaction buffer from New England Biolabs) and XbaI (enzyme and reaction buffer from MBI Fermentas), under conditions described by the suppliers, yielding a fragment of the restriction site polylinker and the linearized plasmid. The linearized plasmid is isolated by standard gel electrophoresis. Single stranded oligonucleotides are synthesized by conventional technology to read 5'GGTCTAGA(VNN)_nCGAGCTC3', in which N represents all four nucleotides, each with equal probability; and V represents C, G, or A, each with equal probability. In a procedure commonly used to generate double stranded oligonucleotides from single stranded templates, a 5' primer and a proof reading

- 56 -

polymerase, such as the Vent polymerase, are used to generate double-stranded oligonucleotides from the single-stranded templates. These double-stranded, highly variable oligonucleotides are then digested with SacI and XbaI, under conditions described by the suppliers, yielding highly variable oligonucleotides with non-complementary sticky ends 5'-CTAG and AGCT-3'. The highly-variable oligonucleotides are then ligated into the linearized plasmid (MBI ligation buffer, conditions described by supplier); they ligate in only one oligo per plasmid, and in only one orientation. Conventional electrophoretic size fractionation is used to isolate ligated plasmids containing an insert. The product of this procedure is a population of plasmids (pEV218I) bearing, at the genetic site corresponding to loop IV of OmpA, a highly variable exogenous sequence free of TAA and TGA stop codons.

To construct a vector population in which each vector harbors between 0 and n variable oligonucleotides at a prescribed site, the following procedure may be used. First, a restriction site that is unique in the vector is located at the prescribed site of insertion, i.e., at a genetic site corresponding to the imprinted domain of the imprinted polypeptide. (If a unique restriction site does not already exist at the desired site, one may be engineered by conventional methods described in the literature, such as oligonucleotide-directed mutagenesis (e.g. Brown, 1992 and Freudl, 1989). If the endonuclease (R) that cleaves at this unique site leaves complementary sticky ends RS and RS', then the variable oligonucleotides generated by the procedures described above are given defined and conserved complementary sticky ends RS and RS'. The vectors bearing the gene for the imprinted polypeptide are cleaved by R. The vectors are then combined with an excess of variable oligonucleotide inserts. Several distinct ligation events can occur. First, the vectors, which have complementary sticky ends RS and RS', can ligate to other vectors or self-ligate, closing without an oligonucleotide insert. Second, the variable oligonucleotides, which also have complementary sticky ends RS and RS' can i) self-ligate as a loop, ii) ligate into a vector with sticky ends RS and RS', iii) ligate to each other, forming a multi-

- 57 -

oligonucleotide concatenation, which itself can perform ligations i, ii, or iii. After ligation, size fractionation can be used to discard all inserts that did not ligate into a vector and all vectors that did not receive inserts. This yields a population of vectors, each of which harbors between 1 and n variable oligonucleotides at the prescribed site in the gene encoding the imprinted polypeptide. A specific example of this method of constructing a vector population in which each vector harbors between 0 and n variable oligonucleotides at a prescribed site may be found in the literature (the construction of such a library of vectors is described in detail in Brown, 1992).

The vectors bearing variable oligonucleotide inserts are used to transform a population of cells that lack a gene for the imprinted polypeptide. Conventional methods of transformation may be used. Transformants are selected by their expression of the vector-borne selectable marker, such as a gene for antibiotic resistance. (Efficient transformation of competent prokaryotic cells achieves up to 10^9 transformants/ml.) The product of transformation is a population of cells, each of which harbors a distinctive oligonucleotide insertion at a genetic site corresponding to a point in the imprinted domain. After transformation, the cell population is amplified by growth in a suitable medium.

In a preferred embodiment of this invention, in which the CISTEM cell population is engineered to display variable oligopeptides at loop IV of *E. coli* OmpA, transformation is performed in the following manner. The variable plasmid population, pEV218I, whose construction was described above, is used to transform *E. coli* strain UH203 (*lac*, *supF*, *ompA*, *recA*, *proAB*, *rpsL*), according to a common method of transformation described in the literature (Freudl, 1989). The UH203 population is grown in Luria broth supplemented with 100 mg Ap/ml, selecting for Ap resistance of transformants.

Expression of the imprinted polypeptide is induced with the compound appropriate to the vector's promoter. Thus, for example, in the preferred embodiment mentioned above, the population of UH203 transformed with pEV218I is induced with 1m IPTG). This yields a population of cells comprising as many

- 58 -

clones as there were transformants in the previous step. Each clone displays a distinctive oligopeptide at the imprinted domain. If the imprinted polypeptide is a glycoprotein or a proteoglycan, glycosylation sites in the imprinted domain bear diverse and unpredictable oligosaccharides. (Commonly practiced procedures for verifying that an extracellular domain-in this case, the imprinted domain-has been glycosylated are described in the literature (Lis et al., 1993).) In short, the CISTEM cell population displays a vast combinatorial library of polypeptides.

In a preferred embodiment, a CISTEM is constructed wherein the imprinted polypeptide is a polytopic membrane protein in which the domain that accepts insertions is directly adjacent to the adhesion domain of a selection agent suitable for use in the CISTEM. This preferred configuration is depicted in figure 5, and several of the polypeptide/selection agent combinations that exhibit this configuration are listed in table 1, and in the literature (Tyler et al., 1996). When these polypeptide/pathogen combinations, or others like them, are used, the CISTEM is assembled with a minimum of genetic manipulation of the imprinted polypeptide; the only necessary manipulation is the introduction of an oligopeptide tag, as well as variable peptide sequences, to the imprinted domain, as described above. However, the list of polypeptides that may be imprinted is by no means limited to polytopic proteins exhibiting adjacent adhesion and imprinted domains. To further expand the diversity of polypeptides that may be imprinted by CISTEMS, two genetic manipulations are employed.

When secreted proteins are imprinted, they are first converted to their glycolipid- or transmembrane-anchored isoforms. Because translocation of secreted proteins and membrane insertion of integral proteins are handled by the same cellular machinery (reviewed in Singer, 1990, and above), many secreted proteins may be converted to membrane-bound isoforms simply by carboxyl-terminal attachment of a hydrophobic transmembrane span or a signal sequence for glycolipid addition. For example, the gene-fragment encoding the 37 carboxyl terminal amino acids of decay accelerating

- 59 -

factor, a eukaryotic glycopospholipid-anchored protein, can be fused to the gene for a secreted protein, such as human growth hormone, in order to target and anchor the protein to the apical surface of human intestinal cell lines. This procedure for converting a secreted protein to its glycolipid-anchored isoform is described in the literature (Cross, 1990). Conversion of secreted polypeptides to isoforms that are anchored by transmembrane spans is also described in the literature. For example, heparin-binding epidermal growth factor-like growth factor (HB-EGF), can be expressed as a longer, membrane-spanning isoform (HB-EGFTMII) comprising an extracellular domain that is structurally identical to the secreted polypeptide (Onu et al., 1994). Other examples of and techniques for the conversion of secreted polypeptides into isoforms anchored by transmembrane spans are described in the literature (Singer, 1990).

A second genetic manipulation of the imprinted polypeptide is performed when the domain that is to be imprinted lacks an adjacent adhesion domain. This occurs e.g. when the polypeptide chosen for imprinting is not polytopic, or when the selection agents that adhere to domains adjacent to the imprinted domain are unknown or poorly suited for use in the CISTEM. In either situation, the sequence for a functional adhesion domain may be fused to the gene for the extra- or intracellular domain, creating a gene for a chimeric polypeptide comprising both the functional adhesion domain and the variable domain. For example, the colicin A adhesion domain from *E. coli* OmpF, when fused to separate domains from another outer membrane protein, OmpC, remains functional as a colicin A receptor (Fourel et al., 1990). A group of eukaryotic viral systems, which are described in the literature (Chu et al., 1995), provide further examples of the insertion of functional adhesion domains. In their application to embodiments of this invention, these systems operate as follows. The antigen-binding site of an antibody (scA) is displayed on the surface of a virus, such as spleen necrosis virus, as described in the literature (Chu et al., 1995). The antigen that is recognized by the scA is then expressed as an exogenous insert in the extracellular domain that will serve

- 60 -

as the adhesion domain of the imprinted polypeptide. (Techniques for inserting exogenous sequences at targeted sites in extra- or intracellular domains are described and referenced extensively above) In this way, a functional adhesion domain for the virus is generated in the desired location, directly adjacent to the imprinted domain.

To briefly summarize the manipulations of the imprinted polypeptide that have been described thus far, a secreted polypeptide may be converted to a membrane-bound isoform, and a functional adhesion domain may be inserted beside the domain that is to be imprinted by this invention. Owing to these manipulations, each of the polypeptide families shown in figure 1 may be imprinted by this invention.

When this invention is used to identify the polypeptides that interact with a given target compound, the CISTEM's imprinted polypeptide may be an outer membrane polypeptide suitable for imprinting. For example, in a CISTEM that identifies polypeptide sequences that interact with the eukaryotic cell-signalling polypeptide Raf, the imprinted polypeptide is *E. coli* OmpA. OmpA extracellular loop IV is imprinted to exhibit high affinity for Raf-RBD. The sequences selected for their affinity for Raf-RBD are then used to screen sequence libraries and cDNA libraries from mammalian cells with active Ras-Raf pathways.

Similarly, when this invention is applied to epitope mapping, the imprinted polypeptide may be an outer membrane polypeptide suitable for imprinting. For example, the CISTEM that imprints OmpA to bind to Raf-RBD is used to map the surface of interaction between the cell-signalling polypeptides Ras and Raf. Sections of the gene encoding Ras are inserted into the genetic site encoding *E. coli* OmpA loop IV. The CISTEM then selects clones that display sections of Ras that exhibit high affinity for Raf. In this way, the CISTEM identifies the epitopes on Ras that bind to Raf.

Certain embodiments of this invention imprint polypeptides to exhibit functional activity. Among these, the preferred embodiments imprint polypeptides that are naturally well-suited to the performance of the functions desired of them. (The structural and functional categories delineated

- 61 -

above provide an outline of the proclivities displayed by distinct families of outer membrane polypeptides). For example, type Ia polypeptides are generally competent for functions involving high-affinity adhesion or specific binding to extracellular targets. They are therefore well-suited to be imprinted for use as highly specific markers or antagonists of small polypeptides, such as small peptide hormones. Type II polypeptides, on the other hand, are often enzymes with extracellular active sites, such as glycosyl transferases. Some type II polypeptides, then, are well suited to be imprinted to exhibit enzymatic activity. When this invention is used to imprint a polypeptide to serve as a functionally active effector, the preferable imprinted polypeptide is usually a membrane-bound isoform of a secreted protein, such as human growth hormone.

Furthermore, in preferred embodiments of this invention, the imprinted polypeptide is a molecule that naturally exhibits functional or structural homology with molecules that perform functions similar or identical to the one desired of the imprinted polypeptide. For example, structural homologues to *E. coli* OmpA have been found in *Haemophilus*, *Shigella*, *Neisseria*, *Bordetella*, and *Streptococcus*. In these virulent bacteria, the OmpA homologues play a role in adhesion to host tissues, and are often recognized by the host's specific immune response. The structural homology between OmpA and these antigenic adhesins makes OmpA well-suited for imprinting for use in applications requiring mimicry of homologous adhesins. Two such applications are competitive inhibition of microbial adhesion, and vaccination, described in greater detail below.

Unlike many other forms of combinatorial chemistry, this invention can screen libraries of polypeptides expressed by eukaryotic cells. This greatly expands the range of molecular properties accessible to this invention, for eukaryotic polypeptides are adorned by saccharides, a second tier of extremely complex, chemically distinctive molecules. The oligo- and polysaccharides adorning the imprinted domain are not varied in the same systematic manner as the underlying polypeptide sequence, because, as was described above, the

- 62 -

saccharide molecule attached to a particular site on a polypeptide is not a simple function of the local amino acid sequence. Instead, the glycosylation at a site is altered in complex and unpredictable ways by changes to the local sequence or to the sequence elsewhere in the polypeptide (Lis et al., 1993). Therefore, when the glycosylation sites within the imprinted domain are preserved while immediately flanking regions are varied by the method described above, a semi-random diversity of sugar structures is appended to a systematic diversity of polypeptide sequences, creating a complex glycoprotein or proteoglycan library.

When a CISTEM's target compound naturally interacts with processed, e.g. glycosylated polypeptides (i.e., in its natural setting, the target compound binds to the processed forms, e.g. to the glycoproteins or proteoglycans), it is preferred that the imprinted polypeptide be processed, especially glycosylated. This requires that the CISTEM cell population be eukaryotic. The adhesions that underpin inflammation, blood clotting, sperm-egg interactions, and intercellular recognition are all mediated by glycoproteins or proteoglycans. Therefore, for the identification of molecules for use as antagonists or agonists in these processes, CISTEMs with eukaryotic cell populations and glycosylated imprinted domains are preferred.

The unpredictable effects of peptide sequence variation upon glycosylation throughout the polypeptide merit a cautionary note. Because changes to a peptide sequence can alter glycosylation on entirely different domains of a polypeptide, the CISTEM's selection (pathogenic) agent must not require a specific glycosylation on its adhesion domain. Otherwise, unexpected changes in the sugars adorning the adhesion domain renders some cells immune to the pathogenic agent even if they do not bind the target compound at their imprinted domains. As was noted above, the selection site, especially an adhesion domain of a pathogenic agent, used in a CISTEM is well-characterized. In most cases, this characterization reveals whether or not the agent requires glycosylations for adhesion. This may be verified, moreover, by removing glycosylations from the receptor and testing for

- 63 -

susceptibility to the agent; methods for removing glycosylations are described in the literature (Lis et al., 1993).

For use as imprinted polypeptides in embodiments of this invention, polypeptides that exhibit structural integrity are preferred. It is particularly preferred that they not only maintain their properties when liberated from the cell surface, but also exhibit stability in the environmental- perhaps *in vivo*-conditions in which the imprinted molecules are ultimately applied. Among prokaryotic polypeptides, β -barrels are exceptionally stable; they can be denatured only by boiling in strong detergents. And in eukaryotic polypeptides, glycosylation often adds structural stability to otherwise malleable protein structures (Lis et al., 1993). In some cases, the structural integrity of imprinted domains may be heightened by genetic manipulations. Such manipulations may be performed before or after imprinting, and they may be used to increase the stability of structures that will be attached to, or liberated from, the membrane.

Genetic manipulations of the imprinted polypeptide may also be used to introduce an element of rational structural design to a CISTEM. Structural analysis of the target compound, its native ligands, or molecules that naturally perform functions similar to that desired of the imprinted polypeptide, may guide the insertion of predetermined motifs into the imprinted polypeptide. A tendency toward the formation of certain structural motifs; such as α -helices, may be incorporated into imprinted domains by inserting variable oligopeptides that are not entirely random, but instead are statistically biased toward sequences that form the desired motifs. Predetermined motifs may also be inserted into the conserved regions of imprinted domains. For example, polyproline or α -helices may be used to extend a variable region away from the cell surface, making it more accessible to a sunken epitope on the target compound. While genetic manipulation informed by structural analysis is by no means necessary to CISTEMs, it may be utilized in combination with CISTEMs to design polypeptides that are well-suited to the performance of desired functions.

- 64 -

In a preferred embodiment, the CISTEM cell population consists of cells that lack a chromosomal copy of the gene for the imprinted polypeptide, but are capable, upon transfection with an episomal or vector-borne copy of this gene, of translating, exporting, and inserting the imprinted polypeptide in the membrane. Thus, for example, in several preferred CISTEMs, which imprint *E. coli* OmpA to bind to any one of several targets, *E. coli* strain UH203 (Freudl, 1989), which is *ompA*⁻, composes the CISTEM cell population. Similarly, in CISTEMs that imprint *E. coli* LamB to bind to a target, *E. coli* strain S1755 (Brown, 1992), which lacks the *lamB* gene, composes the cell population. A eukaryotic example is provided by several preferred eukaryotic CISTEMs, which imprint human ICAM-1 to bind to any one of several targets. In these CISTEMs, COS-7 cells are used; this strain lacks ICAM-1, but is easily transfected with an ICAM-1 gene borne on the CDM8 expression vector (the COS-7/CDM8 expression system is described in Staunton et al., 1990).

The strains that compose CISTEM cell populations are preferably stop-codon suppressors; that is, they recognize one of the stop-codons, such as UAG, as a translatable codon. The reason for this is that any stop-codons recognized by the strain composing the CISTEM cell population must be excluded from the variable oligonucleotides that are inserted into the vector-borne gene for the imprinted polypeptide. (The synthesis of variable oligonucleotides free of certain stop-codons, and their insertion into the gene for the imprinted polypeptide, are described above). Therefore, the synthesis of variable oligonucleotides is simplified when the CISTEM cell population is able to recognize one of the stop-codons. For example, *E. coli* UH203, which is used in several preferred CISTEMs for imprinting OmpA, is *supF*, i.e., capable of translating the UAG stop-codon. Consequently, variable oligonucleotides inserted into the OmpA vector (pEV218), must be free of the stop codons UAA and UGA, but not of UAG. This is easily achieved by the method of synthesis described above. Similarly, S1755, the strains used in CISTEMs imprinting LamB, is *supE44*, i.e., capable of translating UAG.

- 65 -

The medium in which the CISTEM cell population is propagated depends simply on the strain composing the population. For example, *E. coli* UH203 is grown in Luria broth or agar. When a strain contains a plasmid with an antibiotic resistance gene, the appropriate antibiotics are added to the medium or agar. Standard media for all commonly used strains of prokaryotic and eukaryotic cells are described in the literature and familiar to those practiced in the art. The conditions for transformation of CISTEM cells with the vector bearing the imprinted polypeptide also depend upon the cell type, and are described in more detail above, as well as in the literature. Induction of expression of the imprinted polypeptide requires incubation of the cell population with the compound that activates the imprinted polypeptide's inducible promoter. As is described above, however, certain inducible promoters, such as the lac promoter/operator elements, are "leaky", allowing for the appropriate level of expression of the imprinted polypeptide without any induction.

When the CISTEM cell population is incubated with a suspension or solution of the target compound, the medium generally contains only the minimal requisites for cellular sustenance in addition to the target compound. The presence of abundant macromolecules other than the target compound can result in selection for clones whose imprinted domain bind not to the target, but rather to one of the constituents of the medium. Thus, for example, in preferred CISTEMs imprinting OmpA, *E. coli* UH203 are incubated in the presence of any one of several target compounds in ATCC Medium 52. However, in some CISTEMs, Luria broth was used. In several preferred CISTEMs for imprinting ICAM-1, COS cells are incubated with target compound in TC Minimal Medium Eagle or Minimal TC Medium 99. Minimal media for other commonly used prokaryotic and eukaryotic cell lines are described in the literature. The concentration of target compound and duration of incubation are important to CISTEM function, and are discussed in detail below.

When the invention is used for the identification of functionally active polypeptides, such as enzymes or inhibitors of a target compound's biological activity,

- 66 -

incubation with the target compound occurs under chemical and thermal conditions that are as similar as possible to those under which the imprinted polypeptide will ultimately be applied. For example, one preferred eukaryotic CISTEM imprints ICAM-1 to serve as an antagonist to HIVgp120's interaction with the receptor CD-4 (Fantini et al., 1997). In this CISTEM, the cell population is incubated with the target compound, soluble HIVgp120, described above, in temperature and ion concentrations approximating human extracellular fluid (table 2). In the application of the invention to interaction analysis or epitope mapping, the incubation occurs under chemical and thermal conditions that are as similar as possible to those in which the studied interaction naturally takes place.

Many outer membrane polypeptides are ports of adhesion for multiple agents that kill, or retard the multiplication of, the cells they enter. Each of these agents recognizes a distinctive adhesion domain on the polypeptide receptor (figures 2 and 3). The adhesion domains of a wide range of such selection agents, which may be used as selection agents in the present invention, are thoroughly characterized in the literature (table 1, Tyler et al., 1996). Using these characterizations, a selection agent whose adhesion domain resides beside the imprinted domain (as shown in figure 5) is identified. In the event that the polypeptide that is to be imprinted is bitopic, or when a suitable selection agent with an adhesion domain adjacent to the imprinted domain cannot be found, a functional adhesion domain is appended or inserted by genetic manipulations of the imprinted polypeptide; these manipulations are described above.

Once a selection agent with an adhesion domain beside the imprinted domain has been identified, the following procedures are performed to establish and optimize the agent's selection for cells in which the imprinted domain binds the target compound. This procedure serves two functions. First, it establishes that the selection agent significantly impairs cells in which the imprinted domain fails to bind the target compound, while leaving unaffected those cells in which the imprinted domain effectively binds the target. In other words,

- 67 -

the procedure demonstrates that the agent is competent to perform CISTEM selection.

Second, this procedure is used to determine the experimental conditions that will allow reliable CISTEM selection. To adjust CISTEM selection, multiple parameters may be varied, including the following: concentration of the target compound ([T]) with which the CISTEM cell population is incubated; duration of incubation with the target compound; initial multiplicity of infection (MoI) or concentration of the selection agent; duration of incubation with the selection agent; temperature; level of expression of the imprinted polypeptide. Each of these parameters may be optimized by varying the parameter value over a number of CISTEM trials, as will be described here. However, every parameter need not be optimized for every embodiment of this invention. In preferred embodiments of the invention, at least two parameters are optimized by multiple trials, as described here, while other parameters are set by common laboratory procedure or precedents described here. In the optimization described here, concentration of target compound, [T], and initial multiplicity of infection, MoI, are the two parameters that are varied over a number of trials. In the examples below, an optimization is described in which two different parameters - namely, level of expression of the imprinted polypeptide and duration of incubation with the selection agent - are varied.

The procedures for demonstrating that the selection agent is competent to perform CISTEM selection, and for optimizing CISTEM parameters, make use of the cell population whose construction is described above. In this cell population, an oligopeptide tag is exhibited in the imprinted domain. This tag binds with high affinity and specificity to a readily available ligand. (The T7 tag/T7 antibody is a preferred example of a tag/ligand pair; this pair and others are described above and in the literature). In the optimization described here, expression of the imprinted polypeptide bearing the oligopeptide tag is regulated to the relatively low level estimated to be roughly 1000 copies per cell. As was noted above, however, in the optimization of some embodiments of this invention, such as that described in the examples, the

- 68 -

level of expression of the imprinted polypeptide is one of the varied parameters. The method by which regulation of the imprinted polypeptide is achieved depends upon which heterologous promoter controls the gene for the imprinted polypeptide. For example, on pEV218T7, the *ompA* plasmid used in a preferred embodiment of this invention, *ompA* is under the control of the *lac* regulatory elements, which are induced by isopropyl-b-D-thiogalactopyranoside (IPTG). In addition, the *ompA* gene on pEV218T7 comprises a TAG stop codon corresponding to amino acid 7 (Freudl, 1989) of the polypeptide; therefore, when expressed in the amber-suppressing (*supF*) host *E. coli* UH203, *ompA* expression is regulated not only by the level of induction with IPTG, but also by the concentration of the suppressor tRNA, which is expressed from a plasmid. Low level expression of OmpA is achieved with 1.0 mM IPTG induction. On pSB1649, *lamB* is under the control of the *lac* regulatory elements, and expression is regulated as described in the literature (Brown, 1992). Regulation of the expression of ICAM-1 from pCD1.8, the CDM8 vector construct used in a preferred eukaryotic embodiment of this invention, is also described in the literature (Staunton et al., 1990).

Two populations of cells (populations A and B) exhibiting the oligopeptide tag in the imprinted domain are grown in a medium appropriate to the cell type and the expression vector. (For instance, in a particularly preferred embodiment of the invention, *E. coli* UH203 transformed by pEV218T7 are grown at 37°C in Luria broth supplemented with 100µg/ml ampicillin, and 1.0 mM IPTG for induction of pEV218T7 expression; conditions for growth of other cell types are described above and in the literature.) A third population of cells (C) of the same strain as populations A and B, transformed by the same vector as populations A and B, but without the oligopeptide tag, is grown in the same medium. (Thus, for example, in the particularly preferred embodiment mentioned above, *E. coli* UH203 transformed by pEV218, rather than pEV218T7, are grown in the conditions described above.) Populations B and C are each grown to generate three clonal subpopulations of 10^6 cells, B1, B2, B3, and C1, C2, C3. Population A is grown to

- 69 -

generate nine clonal subpopulations, A1, A2...A9, each comprising 10^6 cells.

Populations A1...9 and C1...3 are incubated with the target compound, while populations B1...3 are incubated under identical conditions, but without the target. Optimal conditions and time required for binding between the oligopeptide tag and the target are prescribed by the supplier of the tag/target system or in the literature (Phizicky et al., 1995). In the particularly preferred embodiment, in which the T7 tag is displayed at loop IV of *E. coli* OmpA, and the T7 antibody (Novagen) serves as the target compound, the cell population expressing pEV218T7 is incubated with mAbT7 for approximately 60 minutes. The concentration of target compound with which cells are incubated is systematically varied across populations A1...9 and C1...3 in order to identify the concentration that is optimal for CISTEM selection (table 3). Populations A1...3 are each incubated with the lowest concentration of target compound. This minimum concentration varies, depending upon the properties of the agent and the availability of the target, but is generally approximately 10^3 x concentration of cells x number of imprinted polypeptides per cell. Thus, for example, for a population of 10^6 *E. coli* grown under the conditions described above to express roughly 1000 copies of OmpA per cell, 0.001 mg mAbT7 is used in 1 ml of medium. This is approximately 4×10^3 target molecules (mAbs) per imprinted polypeptide. Populations A4...6 are incubated with a higher concentration of target compound, such as 10x the concentration used with populations A1...3. Populations A7...9 and C1...3 are incubated with the highest concentration of target, such as 100x the concentration used with populations A1...3.

Following incubation, populations A1...9, B1...3, and C1...3 are exposed to the selection agent. The technique of exposure depends upon the nature of the agent, and, for all well-known agents, is described in the literature. Methods of exposing prokaryotic cell populations to bacteriophages are described by Calendar (Calendar, 1988); methods of exposing prokaryotic cell populations to colicins are described by Pattus et al. (Pattus et al., 1990); and methods of exposing

- 70 -

eukaryotic cell populations to viruses are described in the literature (e.g. Staunton et al., 1990, Tyler et al., 1996). Like the concentration of target compound, in the optimization described here, the concentration of agent is varied across the cell populations in order to identify optimal conditions for the CISTEM's selection (table 3). (As was noted above, however, not every parameter is varied in optimizing each embodiment of this invention. In the optimization described below, for example, the concentration of target is not one of the parameters that is varied. Instead, this concentration is fixed, while two other parameters are varied). Populations A1,4,7, B1, and C1 are exposed to the lowest concentration of the selection agent. This concentration will vary from one CISTEM to another, depending upon the properties of the agent, the method and duration of exposure, and the availability (and thus the concentration) of the target compound, but it will generally be roughly the minimum concentration required to affect every cell in a fully susceptible population during a CISTEM run. Thus, for many lytic bacteriophages, such as coliphages K3 or λ , when the method of exposure is classical phage infection (Calendar, 1988), and the duration of incubation with the bacteriophage is 60 minutes at 37°C, the concentration will be 2 - 5 x the cell concentration, or MoI » 2 - 5. Cell populations A2,5,8, B2, and C2 are exposed to a doubled concentration of the agent, and populations A3,6,9, B3, and C3 are exposed to yet another doubling of agent concentration. Table 3 summarizes the variation of target and agent concentrations over the cell populations in this procedure for optimizing CISTEM selection.

The concentrations in this procedure may vary depending on the nature of the selection agent and the method of exposure, but several fundamental aspects of the procedure do not change from one CISTEM to another. First, in the control populations, B and C, the selection agent has unimpeded access to its receptor, and therefore affects every cell. Population B is susceptible to the agent because, although cells bear the tag at the imprinted domain, the population is not exposed to the tag's ligand. Therefore, the selection agent's site, e.g. the adhesion domain, remains exposed and functional.

- 71 -

Population C is susceptible because, although the cell population is exposed to the ligand, the ligand is not bound by the imprinted polypeptide, which lacks the ligand's tag. Because the selection agent's adhesion domain remains exposed and functional in populations B and C alike, the results of exposing these populations to the agent are the same. Thus, in these control trials, exposure to the agent yields the same results for populations B1 and C1; likewise for populations B2 and C2; and for populations B3 and C3. These controls confirm two conclusions critical to the CISTEM: 1. the presence of the target ligand does not affect the selection agent's activity, provided the ligand is not bound to the imprinted polypeptide; 2. the presence of the tag in the imprinted domain does not affect the selection agent's activity, provided the tag is not bound to a target ligand.

Whereas control populations B and C are entirely susceptible to the selection agent, populations A1...9 exhibit a range of resistance, extending even to immunity. In the cells of population A, the imprinted domain harbors the oligopeptide tag, which binds the target ligand during incubation. As a result, the adjacent adhesion domain is occluded from the selection agent, and the cells are protected. Populations A1...9 exhibit a range of resistance because they are exposed to a range of concentrations of target ligand. When the range of concentrations is broad enough, the range of resistance extends from partial susceptibility to complete immunity.

Based on these preliminary trials, the concentration of selection agent that allows for functional CISTEM selection is found as follows. Ubiquitous pathogenic effect--i.e. every cell in a population is affected by the agent--is easily observed when cell populations are exposed to agents according to protocols described in the literature. For example, in exposing prokaryotic populations to virulent lytic phage, ubiquitous pathogenic effect results in confluent lysis of the population, with zero surviving colonies (Calendar, 1988). The lowest concentration of selection agent at which every cell in the control populations (B and C) is affected by the agent is an appropriate level for CISTEM selection. For example, if

- 72 -

some cells from B1 and C1 (table 3) escape the selection agent, but the cells of B2 and C2 are all affected by the agent, then the appropriate concentration of selection agent for the CISTEM is $MoI = 10$. In this example, the minimum concentration required for ubiquitous effect could be more exactly determined by further trials with MoI 's ranging from 5 to 10.

After the CISTEM's concentration of selection agent is found, the optimal concentration of target compound is determined as follows. Each CISTEM has an approximate threshold affinity, $K_a = t$, such that clones whose imprinted domains bind the target with an affinity below this threshold ($K_a < t$) are eliminated from the population by the selection agent, whereas clones whose imprinted domains exhibit an affinity above the threshold ($K_a > t$) are protected from the agent, and thus selected. The threshold for a CISTEM can be set by adjusting the concentration of the target compound. To set the threshold to approximately the affinity that the oligopeptide tag exhibited by populations A and B shows for its ligand, the target concentration used in the CISTEM is the lowest level that provides complete protection to population A_i in the presence of selection agent at the CISTEM's MoI . (Complete immunity to the agent is easily observed when cell populations are exposed to agents according to common protocols described in the literature.) Thus, for example, if the CISTEM's concentration of selection agent, determined as described above, is $MoI = 10$, and population A5 exhibits complete immunity to the selection agent while population A2 is slightly reduced by the agent (table 3), then 10^4 molecules of target compound per imprinted polypeptide is used in a CISTEM to select for cells whose imprinted polypeptides bind the target with an affinity equal to or greater than the affinity exhibited by the oligopeptide tag for its ligand. A slightly lower concentration of target compound, such as 10^3 molecules per imprinted polypeptide, is used to select for cells with imprinted polypeptides that exhibit an affinity for the target that is significantly higher than the affinity exhibited by the oligopeptide tag for its ligand. A higher concentration of target compound, such as 10^5 molecules per

- 73 -

imprinted polypeptide is used to set the CISTEM's threshold affinity well below the affinity exhibited by the tag for its ligand.

The application of CISTEM to cell populations yields cell clones that survive the selection procedure by virtue of the interaction between the polypeptide of interest and the target(s). The sequence of the polypeptide of interest is encoded by the DNA sequence which has been inserted into the display domain. The surviving cell population can be monoclonal or polyclonal. In the latter instance, surviving cells need to be plated out on solid medium plates to recover colonies derived from single cells (state of the art bacterial genetics). Amplification of monoclonal cell populations and the subsequent preparation of their relevant episomal or vector DNA is performed according to standard protocols of recombinant DNA methodology. Since the episomal DNA codes for the display protein and its inserted polypeptide of interest, DNA sequence analysis of the inserted DNA tag (state of the art recombinant DNA technology) provides the DNA sequence of the insert and thereby allows the prediction of the amino acid sequence of the polypeptide of interest which has been expressed and displayed at the surface as integral part of the membrane protein. The inserted DNA is derived from a pool of randomly synthesized oligonucleotides or from DNA fragments derived from natural occurring mRNA (cDNA) or genomic DNA. The selected DNA and cognate polypeptide sequences can be scrutinized to match sequences of known proteins and their cognate genes. Thus, sequence analysis can be employed to identify the polypeptide of interest and to attribute and/or relate its composition to known genes or DNA sequences and their predicted proteins.

Modifications and elaborations of the basic CISTEM achieve several extensions of its capabilities, including the following: i. augmentation of the maximum affinity of selected polypeptides for the target compound; ii. specification of the number of binding sites joining a multivalent imprinted polypeptide to the target compound; iii. selection for imprinted polypeptides that exhibit both high affinity for the target compound and low affinity for a second compound; iv.

- 74 -

selection not only for affinity, but also for functional activity; v. screening the polypeptides expressed from a cDNA library for binding to a target compound; vi. selection for imprinted polypeptides that exhibit affinity for one or more of a set of multiple, distinct target compounds.

To increase the maximum affinity attainable in a CISTEM, iterative CISTEM runs are performed (figure 10). Once a population of clones exhibiting affinity for the target compound has been selected by a CISTEM, the vectors bearing the gene for the imprinted polypeptide are isolated from this population by common techniques appropriate to the vector, and described in the literature. New variable oligonucleotides, such as those described above, are then inserted into the vectors alongside the original variable insertions. This is performed by the same procedure that was used to achieve the original insertion of variable oligonucleotides at a genetic site corresponding to the imprinted domain. (One such procedure is described above). Alternatively, variation is introduced not as an extension of the original oligonucleotide inserts, but rather as an oligomeric rearrangement of the original inserts (the procedure of oligomeric rearrangement of oligonucleotide inserts is described in Brown, 1992) or a limited mutagenesis of the original inserts (techniques for the limited mutation of oligonucleotide inserts are described in Balint et al., 1993; insertion of mutated oligonucleotides is performed according to the procedure described above for insertion of variable oligonucleotides into a targeted site in a vector). In other words, oligopeptide "sequence space" may be searched by a number of strategies, some of which have been applied to screening other combinatorial libraries (Balint et al., 1993, Brown, 1992, Janda, 1994). The vectors bearing fresh variation at a site corresponding to the imprinted domain are then used to transform a population of cells, according to common protocols cited above, creating a new CISTEM cell population. This population is then subjected to yet another round of selection by the CISTEM (figure 10). This cycle of introducing variation, selecting for affinity for the target compound, and allowing selected clones to multiply, is

- 75 -

iterated until binding between the imprinted domain and the target compound attains the desired degree of affinity.

The number of oligopeptide sequences screened by a CISTEM for affinity for the target compound is increased, i.e., sequence space is increased, by performing CISTEMs in parallel, by combining CISTEMs with high-throughput robotics strategies, by performing CISTEMs on a large-scale, or by performing CISTEMs in a continuous-throughput chemostat.

To create imprinted polypeptides that exhibit multivalent binding of the target compound, multiple variable regions may be imprinted by sequential CISTEM runs (figure 11). First, a single region or domain of the imprinted polypeptide is imprinted to bind the target compound. Next, the selected population of cells, which exhibit the previously-imprinted domain, is engineered to exhibit a variable region elsewhere in the imprinted polypeptide-either in an unimprinted region of the previously-imprinted domain, or in an altogether separate domain. This is achieved as follows. First, the vectors bearing the gene for the imprinted polypeptide are isolated from the selected population of cells by common techniques appropriate to the vector and described in the literature. Next, using one of the procedures described above for the insertion of variable oligonucleotides at a specific site in a gene, variable sequences are inserted at a genetic site corresponding to an extracellular domain of the imprinted polypeptide. This site is distinct, however, from the site at which variable oligonucleotides were inserted for the first round of CISTEM selection. The vectors, which now bear selected sequences at one site and variable oligonucleotides at another site, are used to transform a population of cells, creating a new CISTEM cell population. This population is subjected to another round of CISTEM selection. In this round, the concentration of target compound is lower than it was in the previous round, making the threshold affinity for selection, t , higher. The reason for this is that, after the first round of selection, every cell in the CISTEM population binds the target compound, and the CISTEM must now select only cells expressing imprinted polypeptides that bind the target at multiple distinct epitopes. With each new region of the

- 76 -

imprinted polypeptide that is imprinted, a new epitope on the target compound is bound.

Certain embodiments of this invention select for imprinted polypeptides that exhibit high on-rate in their binding with the target compound. In these CISTEMs, the CISTEM cell population is not pre-incubated with the target compound prior to introduction of the selection agent. Rather, the selection agent is introduced simultaneously with the target compound. The concentrations of agent and target compound are varied over multiple trials, as is described in the optimization procedures detailed elsewhere in this document. Concentrations are adjusted such that, upon simultaneous introduction of agent and target to the CISTEM cell population, clones whose imprinted polypeptides bind to the target relatively rapidly are protected, whereas clones whose imprinted polypeptide binds to the target more slowly are eliminated by the selection agent.

A CISTEM that imprints a polypeptide to show high affinity for a target compound but extremely low affinity for a second compound, which is here called the "repelled compound", is constructed as follows. The size of the target compound, the size of the repelled compound, and the distance between the imprinted domain and the adhesion domain are altered, if necessary, to meet the following condition: when the target compound is bound to the imprinted domain, it occludes the adhesion domain, whereas when the repelled compound is bound to the imprinted domain, it is too small to occlude the adhesion domain. If this criterion is not met fortuitously (i.e. if the repelled compound does not happen to be extremely small), then the diameter of the target compound and the distance between the domains may be altered, as follows. The target compound may be enlarged by creating a multimer. (The procedure for creating a multimer depends upon the nature of the target compound, but many techniques for joining multiple copies of a compound are commonly practiced and described in the literature.) The distance between the imprinted domain and the adhesion domain may be increased by moving either domain to a site on the imprinted polypeptide that is topologically farther from the other domain.

- 77 -

Transposing the imprinted domain simply involves inserting variable oligonucleotides, according to the procedure described above, at a genetic site corresponding to a different extracellular domain of the imprinted polypeptide. In many prokaryotic and eukaryotic embodiments of this invention, the adhesion domain may be transposed to a different extracellular domain of the imprinted polypeptide without affecting its function as a receptor for the selection agent. The procedure for transposing an adhesion domain is described above as well as in the literature (Chu et al., 1995). If, as a result of moving the adhesion domain and the imprinted domain farther away from each other, one or more extra- or intracellular domains that are not involved in CISTEM function occur between the two, topologically separated domains that are critical to CISTEM function, the superfluous domains may be replaced by polyproline linkers in order to ensure that extra- or intracellular structures do not impede proper occlusion of the adhesion domain by a target compound bound to the imprinted domain.

Once the three critical distances - the distance between the imprinted domain and the adhesion domain, the diameter of the target compound, and the diameter of the repelled compound - are in accordance with the criterion stated above, the CISTEM cell population is incubated with both the target compound and the repelled compound. Then the CISTEM's selection agent is introduced, and this creates two selective pressures. One pressure favors cells expressing imprinted polypeptides with higher affinity for the target compound (figure 12): a target compound bound to an imprinted domain occludes the adjacent adhesion domain, and thus serves to protect the underlying cell from the selection agent. Among cells that exhibit some affinity for the target compound, a second selective pressure favors cells with lower affinity for the repelled compound; a repelled compound bound to an imprinted domain not only fails to occlude the adjacent adhesion domain, but also prevents binding by a target compound (figure 12). In short, an imprinted polypeptide's affinity for the repelled compound effectively reduces its affinity for the target compound. Thus, selection favors cells

- 78 -

that exhibit imprinted domains with high affinity for the target compound and low affinity for the repelled compound (figure 12).

To select not only for affinity but also for functional activity of imprinted polypeptides, a basic CISTEM may be coupled to a secondary screen for activity. In the first step of coupled selection, a CISTEM selects clones whose imprinted polypeptides bind the target compound in the prescribed fashion. From this population of clones, or from the imprinted domains extracted from such a population, a secondary screen selects the structures whose binding to the target compound exhibits the desired functional activity. Imprinted domains liberated from the cell surface by detergent or by expression as a secreted isomorph (isomorphic forms and their interconversion are described above) may be subjected to any commonly practiced screens or assays for activity or modification of the target compound's activity. For example, one CISTEM imprints *E. coli* OmpA to bind with high affinity to soluble IL-6-Receptor (This target compound is described in more detail above). The imprinted polypeptides, freed from the cell membrane by a conventional gentle detergent, which is known not to denature OmpA, are then subjected to a second screen for their capacity to agonize the cell-bound, functional IL-6-Receptor. This screen for agonistic activity is described in the literature (Hibi et al., 1990). Alternatively, extremely strong selection for functional activity may be exerted upon the CISTEM's selected set of cellular clones, and does not require liberation of the imprinted domains from the cellular membranes. In fact, precisely because they yield replicating cell populations, CISTEMs can be coupled to selection regimes very similar to genetic complementation. In these selection regimes, each cell's supply of a vital compound, such as a nutrient, is dependent upon the imprinted domain's performance of the function desired of it, such as enzymatic activity. In other words, the imprinted domain's functional activity either converts the target compound to a compound vital to the underlying cell, or secures the vital compound for the cell.

- 79 -

In certain embodiments of this invention, the target compound consists of a library of separate and distinct molecules, which is here called „the target library“. A CISTEM in which a target library (rather than a single target compound) is used serves to select interacting pairs in which one member of each selected pair is the oligopeptide displayed at the imprinted domain of a protected clone, and the other member is a target library molecule that binds to this oligopeptide and thus serves to protect the clone. Provided each molecule in the target library is large enough to meet the size requirement for CISTEM function, and is present in the medium at a concentration adequate to confer protection upon cells whose imprinted polypeptides bind to one or more molecules in the target library, the target library may comprise as many distinct molecules as is desired, and may be constructed according to any known procedure for making or isolating libraries of molecules.

In CISTEMs involving target libraries, two strategies are used, if necessary, to increase the concentration of each molecule in the target library to the minimum level required for effective CISTEM selection, i.e., to achieve the concentration required to confer protection from the selection agent upon those cells whose imprinted polypeptides bind to one or more of the molecules in the target library. In one strategy, the CISTEM cell population and the target library are each amplified, divided into fractions, and exposed to one another in microtiter wells. This allows the concentration of target molecules to be sufficiently high to confer protection upon cells whose imprinted polypeptides bind to one or more molecules in the target library.

In the second strategy for increasing the concentration of each molecule in the target library, the target library is secreted by the CISTEM cell population itself. As was described above, libraries of variable oligopeptides can be displayed as fusions to gene protein 3 of *E. coli* filamentous phage (M13, f1, or fd) (filamentous phage-display technology is described in Janda, 1994 and Phizicky et al., 1995). Filamentous phage create relatively avirulent infections in which phage particles are liberated from the host cell by

- 80 -

extrusion, leaving the cell intact. Therefore, a CISTEM cell population that secretes its own target library can be engineered as follows.

As in the basic CISTEM described above, a cell population that lacks a gene for the imprinted polypeptide is transformed with a library of exogenous copies of this gene, in which variable oligonucleotides have been inserted at the genetic site corresponding to the imprinted domain. For example, in the construction of an *E. coli* CISTEM cell population described above, *E. coli* UH203 is transformed with pEV218I, a library of ompA plasmids with variable oligonucleotides at the genetic site corresponding to extracellular loop IV of the membrane protein. The resultant cell population displays a library of exogenous oligopeptides at the imprinted domain. In an extension of the basic CISTEM, this cell population is then transformed with a second extra-chromosomal genetic element that encodes a secreted library of oligopeptides. The OmpA-imprinting cell population, for example, is transfected with a library of filamentous bacteriophage engineered to display exogenous oligopeptides at gene protein 3. (Many such libraries are described in the literature, reviewed in Janda, 1994 and Phizicky et al., 1995.) Each cell in the population now has two extra-chromosomal genetic elements: one encodes OmpA, in which an extracellular loop displays an exogenous oligopeptide; the other encodes a stably extruded M13 phage, in which gp3 displays a second exogenous oligopeptide, distinct from the one displayed on the cell surface (figure 14).

The CISTEM cell population expressing two extra-chromosomal genetic elements is then exposed to the CISTEM's selection agent, following the same protocols for exposure as were described for the basic CISTEM. Upon exposure to the selection agent, a clone in the CISTEM's population is protected only if the target library molecule that is secreted by the clone binds specifically to the clone's imprinted domain. That is, a clone is protected by an autocrine loop. In the specific CISTEM in which *E. coli* displays an OmpA library and secretes an M13 library, K3 serves as the selection agent (figure 13). An *E. coli* clone is protected from K3 only if the

- 81 -

exogenous polypeptide displayed at loop IV of OmpA binds specifically to the M13 that the clone itself secretes. The genes for both members of this interacting pair of oligopeptides are easily recovered from the protected cell. In sum, through autocrine protection of clones, a CISTEM with a target library selects for specifically interacting pairs of oligopeptides, as well as for the genes that encode them.

In another example of autocrine protection, a CISTEM is constructed according to the procedure just described, but a hemolysin library is used in place of the phage M13 library. Hemolysins are secreted bacterial polypeptides that tolerate exogenous insertions. Thus, the hemolysin gene can be used to carry exogenous genes, which encode polypeptides that are then secreted from the cell.

Two possibilities for an intracellular CISTEM are depicted in figure 20. In both possibilities, the protective partner:partner interaction occurs within the cell. Version A shows a typical CISTEM, wherein the two partners are transported to the cell surface. In version B, the positive selection occurs, because the two partners cannot be transported to the cell surface and e.g. accumulate within the cell and/or become substrates for protein degradation pathways.

Certain embodiments of this invention are used to identify antigens associated with tumors. A CISTEM for the identification and isolation of cancer antigens may be constructed as follows: First, a cDNA library is derived from a patient's tumor cells. Protocols for isolation of cells and derivation of the cDNA library are described in the literature. The tumor-derived cDNAs serve as the variable inserts in the construction of the CISTEM cell population: they are inserted into the imprinted polypeptide vector at a genetic site corresponding to the imprinted domain of the polypeptide. The cDNAs are inserted in all three reading frames in order to guarantee that polypeptides associated with the tumor are expressed by the CISTEM cell population. Methods for inserting cDNAs into vectors are available in the art. These vectors are used to transform a population of cells lacking an endogenous copy of the gene for the imprinted

- 82 -

polypeptide. When expression of the vector-borne gene is induced, each clone in the CISTEM cell population displays one tumor-associated polypeptide, fused to the imprinted domain on the cell surface.

The target compound in a CISTEM for the identification of tumor antigens preferably consists of polyclonal antibodies, usually isolated from the same patient whose cells are used as the source for the cDNA library. In other words, the tumor cells that supply cDNA and the antibodies that comprise the target compound are preferably autologous. Methods for the collection of cancer patient serum and the isolation of polyclonal antibodies, which are used as the CISTEM target compound, are described in the literature.

Following incubation of the CISTEM cell population with the target antibodies, the selection agent is added. Any clone in which the imprinted domain displays a tumor-associated polypeptide that binds autologous antibodies has a different status concerning the selection agent than a cell with no such binding. E.g. in case of binding, these cells may be protected from the selection agent. By contrast, any clones in which the imprinted domain displays a tumor polypeptide that is not recognized by autologous antibodies is eliminated. Thus, this CISTEM identifies tumor-associated polypeptides that interact specifically with autologous antibodies. It is shown in the literature that such polypeptides are likely to be immunogenic, eliciting a tumor-specific humoral and cellular immune response.

Certain embodiments of this invention are used to determine whether certain molecules interact specifically with an imprinted polypeptide that has not been altered by the insertion of variable oligopeptides. Any outer membrane polypeptide that is used as a receptor by a pathogenic agent, including polypeptides engineered by the manipulations described above, may be used in these embodiments of the invention. Because the imprinted polypeptide is not engineered to display highly variable oligopeptides, there is no „imprinted domain“, and the selection agent's adhesion domain may be located anywhere on the imprinted polypeptide. As in the basic CISTEM cell populations described above, the

- 83 -

imprinted polypeptide is expressed from an inducible promoter. The cell population, induced to express the imprinted polypeptide, is exposed to a target library, according to one of the methods described above for constructing target libraries and introducing them to CISTEMs. Either microtiter wells or the autocrine loop methods may be used to increase the concentration of target library molecules. The CISTEM's selection agent is then introduced according to procedures described for the basic CISTEM. Cells are protected only if molecules in the target library bind to the imprinted polypeptide, and thus prevent adhesion by the selection agent. Thus, these embodiments of this invention provide an easy and straightforward method for assaying a set of molecules' affinity for the imprinted polypeptide.

Therefore, preferred areas, for which the present method is especially suitable is the discovery and isolation of novel antigens for use in vaccines, tumors, novel targets for therapeutic intervention, novel diagnostic markers, or molecules useful for monitoring treatment efficacy. Such new products, identified, isolated and prepared by the method according to the present invention is also subject matter of the present invention.

According to another aspect, the present invention relates to the cells/the cell population to be used in the present invention with

- a) a membrane polypeptide which has the polypeptide of interest in its extra- or intracellular domain and
- b) a selection site being specific for a selection agent, such that upon binding of a target compound to the polypeptide of interest the selection site of cells which exhibit such a binding of said target compound to said polypeptide of interest is modified in a manner that said selection agent may be used for selecting those cells which exhibit said binding of said target compound to said polypeptide of interest from cells which do not exhibit said binding.

The present invention also relates to a kit comprising the cells/cell population according to the present invention, the target molecule (and/or a library of target compounds) and one or more selection agents. These kits may be provided

- 84 -

specifically for easy definable projects and standardized e.g. for medical purposes, such as treating cancer with immunological methods, treating infectious diseases resulting from infections with highly variable pathogens.

According to another aspect, the present invention is directed to a method for producing a pharmaceutical preparation, which comprises the steps of identifying and isolating a novel pharmaceutically active compound by the CISTEM method according to the present invention, adding pharmaceutically acceptable auxiliary components, such as buffer substances, stabilizers, adjuvants, etc, and completing the pharmaceutical preparation. The novelty of the isolated compound can easily be determined by comparing structural data (amino or nucleic acid sequence, chemical formula, NMR (or other spectroscopy data), etc.) of the isolated compound via databases with the known compounds.

The invention is further explained by the following examples and the accompanying figures without, however, restricting it thereto.

Fig.1 shows type I-IV polypeptides; Fig.2 and 3 show adhesion domains of membrane polypeptides; Fig.4 shows an ICAM-1 protein; Fig.5 shows the components of a CISTEM; Fig.6 shows the binding interactions between the imprinted domains and the target molecules; Fig.7 shows a CISTEM selection using a bacteriophage; Fig.8 shows selection and amplification in a CISTEM; Fig.9 shows OmpA plasmid pEV218; Fig.10 and 11 show an interactive CISTEM run; Fig.12 shows an imprinting example; Fig.13 shows the E.coli/OmpA-CISTEM; Fig.14 shows a CISTEM using bacteriophage K3; Fig.15 and 16 show the preliminary optimization of CISTEM selection by page K3. Fig. 17 to 19 and 21 show the monitoring of the dynamics of the CISTEM selection; Fig. 20 shows two possibilities of intracellular CISTEM

EXAMPLES

i. EXAMPLE 1: AN E. COLI OMPA/COLIPHAGE K3/mAbT7 CISTEM

a) *E. coli* Outer membrane protein A

- 85 -

E. coli Outer membrane protein A is a 325-residue polypeptide comprising an amino-terminal, eight-stranded β -barrel with four extra-cellular domains, three periplasmic loops, and a roughly 150-residue periplasmic carboxyl-terminal moiety (figure 3) (Freudl, 1989, Morona et al., 1985). The β -barrel conformation is extremely stable, and can be denatured only by boiling in strong detergents. Naturally present in approximately $2-3 \times 10^5$ copies per cell, OmpA is among the most abundant proteins in the *E. coli* outer membrane. It is also widely conserved in bacterial evolution.

As was described in the specification, exogenous oligonucleotides, inserted into the ompA gene at sites corresponding to extracellular domains of the polypeptide, are normally translated as part of the polypeptide, and do not disturb the polypeptide's insertion in the outer membrane. Furthermore, the oligopeptides encoded by these exogenous genetic insertions are displayed on the surface of the cell, adopt stable folded structures, and can exhibit enzymatic, agonistic, or antagonistic activities.

b) Construction of a cell population for testing and optimizing CISTEM selection

To construct a cell population for use in testing and optimizing CISTEM selection, an *E. coli* population lacking wild-type OmpA was engineered to express a plasmid-encoded copy of OmpA, in which extracellular loop IV of the polypeptide displayed the T7 tag oligopeptide (Met-Ala-Ser-Met-Thr-Gly-Gly-Gln-Gln-Met-Gly), which binds to the commercially available T7 antibody (Novagene). In pEV218 (Freudl, 1989), the ompA gene, which is under the IPTG-inducible control of the lac promoter, has been engineered to bear a polylinker rich in restriction endonuclease sites at a genetic region corresponding to extracellular loop IV of the polypeptide (figure 9). This plasmid was digested with SacI (enzyme and reaction buffer from New England Biolabs) and XbaI (enzyme and reaction buffer from MBI Fermentas), yielding a fragment of the restriction site polylinker and the linearized plasmid with sticky ends left by the restriction enzymes. The

- 86 -

linearized plasmid with sticky ends was isolated by standard gel electrophoresis. Two single stranded oligonucleotides were synthesized by conventional technology (custom oligonucleotides are available from commercial laboratories, and methods for synthesizing single-stranded oligonucleotides are commonly use and described in the literature) to read as follows:

Oligonucleotide 1:

5' - CG[ATGGCTAGCATGACTGGTGGACAGCAAATGGGT] - 3'

Oligonucleotide 2:

5' - CTAG[ACCCATTTGCTGTCCACCAGTCATGCTAGCCAT]CGAGCT - 3'

The sequence in brackets composes the T7 tag gene sequence. These two single-stranded oligonucleotides were annealed in Klenow reaction buffer (MBI catalogue), forming the double stranded T7-tag gene with sticky ends complementary to the sticky ends on pEV218, linearized by SacI and XbaI as described in the previous paragraph. When the double-stranded oligonucleotide was ligated (MBI Fermentas ligation buffer) into the linearized plasmid, the T7 tag gene was inserted, in frame, into the ompA gene on pEV218. The new construct was designated pEV218T7.

The plasmid pEV218T7 was used to transform a population of the strain UH203 (lac, supF, ompA, recA, proAB, rpsL) (Freudl, 1989), according to common methods of transformation described in detail above.

c) Preliminary optimization of CISTEM selection by coliphage K3

Coliphage K3 invades *E. coli* by binding to loops I, II, and III of OmpA. Loop IV is not used by the phage, and, indeed, exogenous oligopeptides at loopIV do not affect K3 susceptibility of the cell (the K3 adhesion domains are shown in figure 3; bacteriophages are reviewed briefly in the specification, and bacteriophage K3, as well as its receptor requirements, are described in Freudl, 1989 and Morona et al., 1985). The following procedure was performed to fulfill two

- 87 -

objectives: a) to conclusively demonstrate that cells whose imprinted polypeptides bind the target compound are protected from K3 attack, whereas cells whose imprinted polypeptides fail to bind the target are eliminated by the phage, and b) to achieve preliminary optimization of the duration of incubation of the CISTEM cell population with phage K3. The principle of this experiment is illustrated in figure 14.

E. coli MC4100 (pEGFP-N1), a kanamycin-resistant, ompA wild-type strain, which will here be referred to as U (for „Untagged“), was cultivated overnight in Luria broth (LB) supplemented with 25 µg kanamycin/ml. *E. coli* UH203 (pEV218T7), which will here be referred to as T (for „Tagged“), is an ampicillin-resistant, ompA⁻ strain that has been transformed, as described in the specification, by a plasmid encoding OmpA bearing the T7 tag sequence at the fourth extracellular loop of the polypeptide. This strain was cultivated overnight in LB supplemented with 100 µg/ml ampicillin/ml and 5 mM IPTG for induction of pEV218T7 expression. Each overnight culture was diluted with LB to OD₆₀₀ = 0.001. To create each of two mixed populations, 50 µl aliquots of the U culture were combined with 50 µl aliquots of the T culture. One mixed population was incubated for 20 min. at 37°C with 20 µl PBS alone, while the other incubated under the same conditions with 20 µl PBS plus 0.1 µg anti-T7 tag mAb (Novagene mAbT7). (In table 4, the presence or absence of mAbT7 is indicated by a + or -.) Phage K3 diluted in 1 ml LB were then introduced to each mixed population at a Multiplicity of Infection (MoI) = 10. (Phage stock was prepared by the conventional method for preparation of T-even phages (Calendar, 1988).) Sample aliquots of 100 µl were taken from each population at the appointed times shown in table 4, namely, immediately preceding introduction of phage (shown in the table as T = 0, Phage -), immediately following introduction of phage (shown in the table as T = 0, Phage +), and at 2 min., 10 min., 30 min., 60 min., 90 min., 120 min., 180 min., and 240 min. after the introduction of phage. Aliquots were plated on LB plates supplemented with 100 µg ampicillin/ml and on LB plates supplemented with 25 µg kanamycin/ml. Colonies were counted after incubation of the plates overnight at 37°C.

- 88 -

The results of this experiment demonstrated that interaction between mAbT7 and the T7 tag at loopIV of OmpA protected cells from phage K3. This conclusion is summarized most concisely in table 5. Cells exhibiting the T7 tag at loopIV of OmpA were protected from phage K3 infection, and this protection was dependent upon incubation of the cells with mAbT7. Cells that expressed wild-type OmpA, rather than the OmpA-T7 tag fusion were susceptible to phage K3, irrespective of the presence or absence of mAbT7. Thus, cells whose imprinted polypeptides bind the target compound are protected from K3 attack, whereas cells whose imprinted polypeptides fail to bind the target are eliminated by the phage.

This experiment also serves in the preliminary optimization of CISTEM selection by phage K3. As is shown in table 4 and figure 15, in the mixed population of U and T cells that were incubated with mAbT7, U cells were completely eliminated from the population after 30 min. incubation with phage K3. T cells, by contrast, maintained a roughly stable population well beyond 30 min. of incubation with the phage. This suggested that when starting with MoI = 10, the minimum time period for phage K3 to eliminate all cells that failed to bind the target would be 30 - 60 min.

The initial precipitous drop in T cells followed by abrupt stabilization (see figure 15) suggested that the antibody may not yet have come to equilibrium in binding the T7 tag at the time the phage were introduced. Pre-incubation with the antibody was 20 min. in duration, and complete protection of T cells was apparently achieved at approximately 30 min. after the introduction of phage, suggesting that the time required for the antibody to come to equilibrium in binding the tag was approximately 50 min.. This suggestion was confirmed by the following preliminary optimization of the duration of the cell population's incubation with the target.

Strains U and T were cultivated overnight under the conditions described above. Each of five mixed populations of U and T cells was created by taking a 50µl aliquot of the U culture (OD₆₀₀ = 0.001) and combining it with a 50µl aliquot of the T culture (OD₆₀₀ = 0.001). As is shown in table 6 and

- 89 -

figure 16, each mixed population was incubated at 37°C with 7nM mAbT7 diluted in PBS for a different time period: 1 min.; 10 min.; 20 min; 30 min; or 60 min.. Phage K3 stock, diluted with LB to 1 ml and MoI = 10, was added to each mixed cell population immediately following the population's incubation with the antibody. Each population was incubated with phage for 30 min. Aliquots were then taken and their composition of U and T cells determined as described above. As is shown in table 6, cells exhibiting wildtype OmpA remained unprotected from phage irrespective of the duration of incubation with mAbT7. By contrast, cells exhibiting the T7 tag at loop IV of OmpA received increasing protection from the antibody as the duration of incubation was lengthened. As was expected from the results the preliminary optimization of phage K3 incubation (described above) the antibody conferred complete protection upon T cells only after an incubation time between 30 and 60 min.

d) Initial test of CISTEM selection

The preliminary, single-parameter optimizations of incubation time with mAbT7 and the duration of K3 selection allowed an initial test of the CISTEM's capacity to select rare cells whose imprinted polypeptides bind the target from a background of cells whose imprinted polypeptides fail to bind the target. To demonstrate the selection of rare cells, the following procedure was performed.

E. coli strains U (Untagged) and T (Tagged) are described above. U was cultivated overnight in LB supplemented with 25µg kanamycin/ml. T was cultivated overnight in LB supplemented with 100µg ampicillin/ml and 1mM IPTG to induce expression of pEV218T7, the plasmid encoding OmpA with the T7 tag at loopIV of the polypeptide. Each overnight culture was diluted with LB to OD₆₀₀ = 0.005. A 50µl aliquot of the U culture was combined with a 50µl aliquot of the T culture to create a mixed population. As is shown in table 7, although the mixed population was created with equal sized aliquots of U and T cultures, it was composed predominantly of U cells. The reason for this was probably that the very low level of OmpA

- 90 -

expression induced in the T strain (only 1mM IPTG was used) made this strain significantly less viable than the U strain under standard conditions of cultivation. The mixed population was incubated at 37C with 68nM mAbT7 (described above) diluted in PBS for 60 min.. Following incubation, 100 μ l aliquots of the cell population were plated on LB plates supplemented with 25 μ g kanamycin/ml or 100 μ g ampicillin/ml. Colonies were counted after overnight growth, showing that the population consisted of approximately 1 T cell for every 120 U cells (table 7). Phage K3 stock diluted in LB was then added to the cell population at MoI = 2, bringing the total volume to 1 ml. The cells were incubated with phage for 120 min. at 37C. Following incubation with phage, 100 μ l aliquots of the cell population were plated on LB plates supplemented with 25 μ g kanamycin/ml or 100 μ g ampicillin/ml. Colonies were counted after overnight growth.

Control experiments were performed according to the same protocol, except that PBS was added without mAbT7 and/or LB was added without phage K3, as shown in table 7. The results in table 7 demonstrated that rare cells (1 in > 100) exhibiting an imprinted domain that bound the target were completely protected from phage K3 while all cells that failed to bind the target were quickly eliminated from the population.

A systematic test of the CISTEM's sensitivity in selecting rare cells that bind the target from a vast excess of cells that fail to bind the target is performed as follows. Overnight cultures of *E. coli* strains U (Untagged) and T (Tagged) are described above. 50 μ l - 1 ml aliquots of the U and T overnight cultures are combined to create four populations, each representing a distinct ratio of U cells to T cells: 10^8 U to 10^4 T; 10^8 U to 10^3 T; 10^8 U to 10^2 T; and 10^9 U to 10^2 T. Each population is incubated for 60 min. at 37C with 10nM mAbT7. To monitor the population composition (i.e. the numbers of U and T cells), every 30 min. (i.e., at 30 min. and 60 min.) two aliquots of 5 μ l are taken, one aliquot is plated on LB plates supplemented with 100 μ g Ap/ml, and the other aliquot is plated on LB supplemented with 25 μ g Km/ml. The plates are then incubated overnight and colonies

- 91 -

counted to determine population composition. Immediately after removal of aliquots at 60 min., bacteriophage K3 (described above) are introduced to the cell population at MoI = 2. The admixture is incubated at 37°C for 180 min.. Again, to monitor the population composition, a sample aliquot of 5 µl is taken by micropipette every 20 min., plated, grown, and U and T cells counted as described above. After incubation with phage, a sample of approximately 1000 cells is removed, plated, grown, and its composition determined as described above.

Demonstration of the CISTEM's capacity to select rare cells that bind the target compound from a background of at least 100 times as many cells that fail to bind the target permits a final, two-parameter optimization of CISTEM performance. Each of five populations of T is induced for 15 min. by a distinct amount of IPTG in the medium: 0 mM; 0.25 mM; 0.5 mM; 1.0 mM; or 2.0 mM. Five mixed populations of 10^5 U and 10^3 T, each representing a different level of induction of T cells, are created by combining aliquots, as described above. 10 nM mAbT7 is added to each population, and the populations are incubated for 60 min. at 37°C. During the 60 min. incubation, the population composition is monitored, as described above, by taking sample aliquots of 5 µl every 20 min., plating, growing, and counting the numbers of U and T cells. After the removal of an aliquot at 60 min., coliphage K3 is introduced at an MoI of 2. Over the next 180 min., the populations are monitored by 5 µl aliquot sampling every 20 min.. The [IPTG] and duration of incubation with K3 that result in the elimination of all U cells, with concomitant elimination of fewest T cells, are optimal parameter values for use in the CISTEM.

The optimization trials described here provide parameter-values ([IPTG], duration of incubation with K3, [mAbT7], duration of incubation with mAbT7) that allow the CISTEM to destroy all cells that fail to bind the target, while eliminating as few high-affinity cells as possible.

e) Construction of an OmpA-imprinting CISTEM cell population

- 92 -

OmpA plasmid pEV218 (described above, shown in figure 9, construction described in Freudl, 1989), bears a polylinker rich in restriction endonuclease sites at a genetic region corresponding to extracellular loopIV of the polypeptide. This plasmid is digested with SacI (enzyme and reaction buffer from New England Biolabs) and XbaI (enzyme and reaction buffer from MBI Fermentas), under conditions described by the suppliers, yielding a fragment of the restriction site polylinker and the linearized plasmid. The linearized plasmid is isolated by standard gel electrophoresis. 10^6 distinct, single-stranded oligonucleotides are synthesized by conventional technology (synthesis of variable oligonucleotides is described above) to read as follows:

5'NNGGTCTAGA(VNN)20CGAGCTCNN3'

in which N represents all four nucleotides, each with equal probability; and V represents C, G, or A, each with equal probability. (*E. coli* strain UH203 is *supF*, capable of translating the UAG stop-codon. Therefore, variable oligonucleotides inserted into the OmpA vector pEV218 must be free of the stop codons UAA and UGA, but not of UAG.) Using a standard protocol described in the literature (Brown, 1992), a 5' primer and a proof reading polymerase (such as the Vent polymerase) are used to generate double-stranded oligonucleotides from the single-stranded templates. These double-stranded, highly variable oligonucleotides are then digested with SacI and XbaI, under conditions described by the suppliers (New England Biolabs and MBI Fermentas, respectively), yielding highly variable oligonucleotides with non-complementary sticky ends. The highly-variable oligonucleotides are then ligated (MBI ligation buffer, conditions described by supplier) into pEV218, linearized as described in the previous paragraph. Because the SacI and XbaI sticky ends are non-complementary, they ligate in only one oligonucleotide per plasmid, and in only one orientation. Conventional electrophoretic size fractionation is used to isolate ligated plasmids containing an insert. The product of this procedure is a population of plasmids (collectively

- 93 -

designated pEV218I) bearing, at the genetic site corresponding to loop IV of OmpA, a highly variable exogenous sequence free of UAA and UGA stop codons.

The variable plasmid population, pEV218I, is used to transform *E. coli* strain UH203 (*lac*, *supF*, *ompA*, *recA*, *proAB*, *rpsL*) (this strain, its propagation and its transformation, are described more completely above, and in Freudl, 1989). The UH203 population is grown in Luria broth supplemented with 100 µg Ap/ml, yielding a population of 10^8 cells comprising 10^6 clones, each one of which carries a distinct member of the pEV218I population.

f) CISTEM-selection for OmpA exhibiting high-affinity for mAbT7

The UH203 (pEV218I) population of 10^8 cells plus approximately 100 UH203 (pEV218T7) cells is incubated at 37°C for 15 min. in 100ml LB medium supplemented with 100 µg Ap/ml and IPTG between 0.25 and 2.0 mM, as determined in the optimization trial described above. This yields a population of 10^8 clones, comprising 10^6 distinct clones, each of which displays a distinct, exogenous, oligopeptide sequence at loop IV of OmpA. One of these clones displays the T7 tag sequence. After the initial 15 min. induction of pEV218I/pEV218T7 expression, 10nM mAbT7 is added to the 100ml of cells. The cells are incubated with the mAb for 60 min. at 37°C. Coliphage K3 is then introduced at MoI = 2 (the procedure for growth and handling of K3 is described above, and in Freudl, 1989, 126, and Morona et al., 1985). Incubation with K3 proceeds for the minimum amount of time required, as determined in the optimization trial described above. Following incubation with phage, the cell population is centrifuged to separate cells from phage and the products of lysis, and cells are resuspended in 100ml NB, without 100mM NaCl supplement, to guarantee complete cessation of infection by K3. This cell population represents the set of clones selected by the CISTEM for binding to the target compound mAbT7.

- 94 -

To amplify the numbers of selected clones, the selected cell population is grown to 10^6 cells at 37°C in 100ml NB. Expression of OmpA is then heavily induced by 2.5 mM IPTG for 15 min.. Following induction, 10^4 cells from the selected population are plated onto NB agar containing 0.5% sodium deoxycholate. (Sodium deoxycholate prevents growth of cells that fail to insert OmpA in the outer membrane; such cells are "false positives", selected by the CISTEM simply because they display no OmpA, and not because of their capacity to bind mAbT7 at loop IV of OmpA. The sodium deoxycholate supplement also prevents K3 infection, providing a further guarantee against continuing infection.) Plated cells are incubated at 37°C to grow colonies, which are composed of cells in which loop IV of OmpA contains an exogenous oligopeptide with high affinity for the target compound, mAbT7.

ii. **EXAMPLE 2: AN *E. COLI* OMPA/COLIPHAGE K3/NEISSERIA MENINGITIDIS ANTIBODY CISTEM**

a) ***E. coli* Outer membrane protein A and *Neisseria* Opa proteins**

Several recent findings show that one important OmpA function explaining such high levels of expression and evolutionary conservation of the polypeptide is adhesion to and invasion of host tissues. It has recently been demonstrated that *E. coli* OmpA contributes significantly to invasion of brain microvascular endothelial cells a critical step in pathogenesis of neonatal *E. coli* meningitis. Similarly, OmpA homologs are directly involved in microbial adherence, invasion of host cells, and pathogenesis by *Haemophilus influenzae*, *Neisseria meningitidis*, *Neisseria gonorrhoeae*, and other virulent microbes.

The structural homology between OmpA and a wide range of antigenic adhesins makes OmpA well-suited for imprinting for use in applications requiring mimicry of homologous adhesins. One such application is vaccination. When a polypeptide (in this case, OmpA) is imprinted to bind to antibodies composing the sera raised against an antigen, the imprinted polypeptide

- 95 -

may elicit an immune response very similar to the one provoked by the antigen (Folgori et al., 1994). Antibodies suitable for use as the target compound in a CISTEM for imprinting *E. coli* OmpA to mimic *Neisseria meningitidis* Opa proteins (Olyhoek et al., 1991) include mAb B33, which reacts with all Opa proteins, mAbs P322 and P514, which react with a broad subset of Opa proteins, and mAb B306, mAb A222/5b (Olyhoek et al., 1991), and others described in the literature.

b) CISTEM-selection for OmpA exhibiting high-affinity for antibodies against

***Neisseria meningitidis* Opa proteins**

The construction of a cell population for testing and optimizing CISTEM selection, the optimization and demonstration of CISTEM function, and the construction of an OmpA-imprinting CISTEM cell population are performed exactly as described above: this CISTEM, like the one described in Example 1, uses phage K3 to select from a cell population displaying variable inserts 20 amino acids in length at loop IV of OmpA. However, whereas Example 1 imprints OmpA to bind to mAbT7, this CISTEM imprints OmpA to bind to antibodies against *Neisseria meningitidis* Opa proteins. Therefore, where the protocol described in Example 1, part f, calls for 10 nM mAbT7, this CISTEM calls for 10 nM of each of the following antibodies: B33; P322; P514; B306; and A222/5b. Aside from this substitution, the protocol for imprinting OmpA to bind anti-Opa antibodies is as described in Example 1, part f.

iii. EXAMPLE 3: MONITORING THE DYNAMICS OF CISTEM SELECTION

To monitor the process by which cells that fail to bind the target compound are eliminated, while those cells that bind the target compound specifically are selected, the following experiment was performed. *E. coli* strains U (Untagged) and T (Tagged) are described above (in Example 1). U was cultivated overnight in LB supplemented with 25µg kanamycin/ml. T was cultivated overnight in LB supplemented with 100µg

- 96 -

ampicillin/ml and 2mM IPTG to induce expression of pEV218T7, the plasmid encoding OmpA with the T7 tag at extracellular loopIV of the mature polypeptide. Each overnight culture was diluted with LB to OD600 = 0.005. To create each of three mixed populations, aliquots of the two cultures were combined to compose a population in which the ratio of T cells to U cells was approximately 1 to 200. Sample aliquots were taken from the populations and plated on LB plates supplemented with 100µg ampicillin/ml, and on LB plates supplemented with 25µg kanamycin/ml. After overnight incubation, colonies were counted, showing that the ratio of Tagged to Untagged cells was approximately 1T to 190U (table 8, t= -60). Two of the three mixed populations were incubated for 60 min. at 37C without mAb, whereas the third population was incubated under the same conditions with 68nM mAbT7 (described above). (In table 8, the presence or absence of mAbT7 is indicated by a + or a -, respectively.) After one hour of incubation (i.e., at time t=0 in table A1), Phage K3 was introduced to two of the three mixed populations at a Multiplicity of Infection of 6. (In table 8, the presence or absence of phage K3 is indicated by a + or -, respectively.) Over the next 180 min., sample aliquots were taken from the populations at the times indicated in table 8, namely, immediately preceding the addition of phage, and then every 15 minutes thereafter. Aliquots were used to estimate phage titer, and were plated on the appropriate antibiotics to determine numbers of U and T cells.

Figure 17 shows the dynamics of phage titer and MoI over the course of CISTEM selection. MoI climbs exponentially and reaches a maximum near 100,000. Figure 18 shows the dynamics of cell growth and death during CISTEM selection. The rapid exponential growth of phage results in a monotonic decline of U cells, irrespective of the presence or absence of mAbT7. By contrast, in the presence of mAbT7, T cells are protected from the extremely high MoI, and consequently they expand in number. Thus, specific interaction between the antibody and the epitope displayed at OmpA loopIV resulted in selection of cells exhibiting the epitope, while cells that did not exhibit the epitope were eliminated. This result is summarized in figure 19.

- 97 -

iv: EXAMPLE 4: REDUCTION OF BACKGROUND THROUGH THE USE OF
HELPER CELLS

The term "background" is used to refer to cells that fail to bind the target compound but nonetheless remain in the selected population. In general, it is desirable to minimize the number of background cells, such that CISTEM reliably eliminates cells whose imprinted domains fail to bind the target compound, while selecting those cells whose imprinted domains do bind the target compound. The following experiment illustrates one strategy for reducing background. The CISTEM cell population is supplemented with "helper cells", the function of which is propagation and amplification of the selective agent. This amplification of the selective agent increases the intensity of selection, and thereby reduces residual background cells. The helper cells lack a selectable marker that is present in the rest of the CISTEM cell population; this permits efficient elimination of the helper cells after they have served their purpose of amplifying the selective agent.

UH203(pEV218T7), the "T" strain, was cultivated overnight in LB supplemented with 100µg ampicillin/ml and 1mM IPTG to induce expression of the ompA plasmid. As a background strain that does not exhibit the T7 epitope at loopIV of OmpA, UHF1(pEV218) was used. UHF1 is virtually isogenic with UH203, except that it carries a gene for Chloramphenicol resistance. The plasmid pEV218 is identical to pEV218T7, except that it carries a restriction-enzyme site polylinker in place of the gene for the T7 tag. UHF1(pEV218), the "U" strain, was cultivated overnight in LB supplemented with 100µg ampicillin/ml and 1mM IPTG to induce expression of pEV218. To create mixed populations, aliquots of the "T" and "U" strains were combined to compose populations with a ratio of T to U cells of approximately 1 to 100 (table 9). As helper cells, MC4100 (pEGFPN1) were used. 100µl of MC4100 (pEGFPN1) at OD₆₀₀ = 0.5 were added to two mixed populations. Two other mixed populations received no helper cells. One population with helper cells, and one without helper cells, were incubated for 1 hour at 37C with 680nM mAbT7, while the other two populations were incubated under the same conditions without antibodies. After incubation

- 98 -

(i.e., at time $t=0$ on table 9), phage K3 were added to all populations at an MoI = 2. All populations were then incubated for two hours. Aliquots were taken and plated on agar supplemented with ampicillin alone, or with ampicillin plus chloramphenicol. The numbers of U and T cells in the population at the end of CISTEM selection are shown in table B1. The background proportion, defined as U cells/(U cells + T cells), is reduced from 34% in the absence of helper cells to 16% with the addition of 100 μ l of helper cells at OD₆₀₀. This experiment thus illustrates the principle of helper cells: they are added to CISTEM cell population to modify the activity of the selective agent.

v. EXAMPLE 5: THE SENSITIVITY OF CISTEM

The following experiment was performed to demonstrate the sensitivity of CISTEM selection. The *E. coli* strains UH203(pEV218T7), here called "T" cells, and UHF1(pEV218), here called "U" cells, were cultured overnight with 1mM IPTG to induce expression of the ompA plasmids. To create mixed populations, aliquots of the T culture were diluted 10, 100, 1000, or 10,000 times, and then combined with undiluted aliquots of the U culture to compose populations with T to U ratios as shown in table 10. Helper cells were added as described in the previous example. One set of five mixed populations, representing each of the serial dilutions of T cells (see table 10), was incubated for 1 hour at 37C with 340 nM mAbT7. Two replicate sets of five mixed populations (a total of 10 populations, comprising two of each of the serial dilutions of T cells) were incubated under the same conditions without the antibodies. Immediately following incubation, phage K3 was added to the set of five mixed populations that had been incubated with the antibodies, as well as to one of the two sets of five mixed populations that were not incubated with the antibodies. In table 10, the presence of antibodies or phage is indicated by a +, whereas the absence of antibodies or phage is indicated by a -. After two hours of incubation, aliquots from each population were plated on agar containing the appropriate antibiotics, as described above. After overnight culture,

- 99 -

colonies were counted to determine the composition of the CISTEM cell populations.

Table 10, shows not only the final numbers of cells, but also several indices of CISTEM's performance in selecting T cells from a background of U cells. Enrichment, an index of the selection system's capacity for amplifying clones that bind the target compound while eliminating those that do not, is defined as $(T \text{ cells}/U \text{ cells})_{\text{after selection}} / (T \text{ cells}/U \text{ cells})_{\text{before selection}}$. Yield, an index of the proportion of all cells that bind the target compound that are successfully selected by the CISTEM, is estimated by comparing the number of T cells in the K3- populations with the number of T cells in the K3+, mAb+ populations. Yield is defined as $(T \text{ cells})_{K3+, mAb+} / (T \text{ cells})_{K3-}$.

vi.: EXAMPLE 6: MAPPING THE EPITOPE RECOGNIZED BY mAbT7

One application of CISTEM is in mapping the epitopes involved in protein-protein interactions. To illustrate this application, CISTEM was used to select cells displaying the phage T7 polymerase epitope that is recognized by the T7-polymerase specific mAbT7 (described above). The plasmid pET17xb (Novagene) contains the entire gene for phage T7 polymerase. This plasmid was digested with micrococcal nuclease, according to protocols provided by the supplier, to yield fragments of a range of sizes from approximately 50 basepairs to 300bp. The band corresponding to fragments between 60 and 75 bp in length was excized with a razorblade and the fragments were resuspended. The fragments were treated with S1 nuclease (conditions according to supplier) to remove the 3' phosphates left by micrococcal nuclease. The fragments were then phosphorylated by T4 polynucleotidkinase. The ompA plasmid pEV218 (described above) was cleaved by the restriction enzymes Ecl136II and SalI (conditions of digestion according to supplier), at unique sites in the genetic region corresponding to the fourth extracellular loop of the mature polypeptide. The linearized vector was dephosphorylated by calf intestinal phosphatase. Combining 100ng linearized vector with 200ng fragments, the fragments were ligated into the linearized vector by standard blunt-end ligation using T4 ligase. PEG was added to the ligation buffer to favor circularization.

- 100 -

Circularization of the vector without insertion of a fragments creates a SalI site. Therefore, the products of ligation were treated with SalI to linearize vectors that did not contain a fragment. The circularized vectors containing fragments were isolated by electrophoresis, and used to transform the highly competent, commercially available strain XL-1 blue. Cells were plated on LB supplemented with 100µg/ml ampicillin to select for transformants. Colonies were grown in 250ml LB supplemented with ampicillin, as well as glucose for catabolic repression of low-level expression of the ompA plasmid. Plasmid DNA was extracted using the Qiagene kit. 7µg of plasmid DNA was cleaved by SalI and used for transformation of UH203 (an ompA deficient strain described above) by electroporation and CaMg transformation. Cells were grown on LB supplemented with 100µg ampicillin to select for transformants. This population of transformants will be referred to as UH203(pEV218T7L).

The UH203(pEV218T7L) population was grown in overnight culture with 1mM IPTG to induce expression of the ompA plasmid. The cells were incubated for one hour with 680nM mAbT7, allowing for binding between the monoclonal antibody and the epitope presented at OmpA loopIV by certain members of the cell population. Following incubation, phage K3 were added at an MoI = 2, and the cells were incubated with phage for two hours. The cells were then plated, and the results of selection (shown in fig.21) are summarized in table 11. 25 colonies were sampled, their plasmids were isolated as above, and the plasmids were subjected to cleavage by the restriction endonuclease NheI. The genetic sequence encoding the epitope recognized by mAbT7 (i.e., the T7 tag) contains a NheI site; no other NheI site appears in pEV218. Therefore, only plasmids comprising the sequence encoding the epitope of interest are vulnerable to NheI. 24 of the 25 clones sampled were vulnerable to NheI.

TABLES

Table 1: Each polypeptide (column A) is listed with the selection agent (column B) that could be used in a CISTEM to imprint the polypeptide; the respective domains that compose the agent's adhesion domains (column C; numbering begins at the amino terminal); a domain that is interacting (e.g. for sterical reasons) with the adhesion domain and could be imprinted to bind a target compound (column D)

| A: Polypeptide | B: Selection Agent | C: Adhesion Domain | D: Proposed Imprinted Domain |
|-------------------------------------|--------------------------|-------------------------------|------------------------------------|
| OmpF | Colicin A | extracellular loops 1-5 | extracellular loops 6 and 7 |
| | Colicin N | extracellular loops 1, 3-6 | extracellular loops 2 and 7 |
| OmpA | see figure 3 | see figure 3 | any of 4 loops |
| ICAM-1 | rhinovirus | Ig domain 1 | Ig domains 2 and 3 |
| MHV | mouse | domain 1 | all other |
| receptor: 120 kD glycoprotein | hepatitis virus | | domains |

Table 2

| Component | Intracellular Conc. (mM) | Extracellular Conc. (mM) |
|------------------|------------------------------|------------------------------|
| <u>Cations</u> | | |
| Na ⁺ | 5-15 | 145 |
| K ⁺ | 140 | 5 |
| Mg ²⁺ | 30 | 1-2 |
| Ca ²⁺ | 1-2 | 2.5-5 |
| H ⁺ | 4x10 ⁻⁵ or pH 7.4 | 4x10 ⁻⁵ or pH 7.4 |
| <u>Anions</u> | | |
| Cl ⁻ | 4 | 110 |

Table 3: Serial dilutions of selection agent and target concentrations in preliminary optimization of CISTEM selection. The cell populations listed in the table are described in the text.

Mol or [Selection Agent]/[Cells]

| | | | | |
|---|-----------------|---------------------------------|---------------------------------|---------------------------------|
| <div style="border: 1px solid black; padding: 5px; text-align: center;"> <u>[Target Compound]</u> [Imprinted Polypeptide] </div> | 0 | 5 | 10 | 20 |
| | 0 | B ₁ | B ₂ | B ₃ |
| | 10 ³ | A ₁ | A ₂ | A ₃ |
| | 10 ⁴ | A ₄ | A ₅ | A ₆ |
| | 10 ⁵ | C ₁ , A ₇ | C ₂ , A ₈ | C ₃ , A ₉ |

Table 4

| Time [min] | Phage | Antibody | Selection | Selected | Number of colonies |
|------------|-------|----------|-----------|----------|--------------------|
| 0 | - | - | LB | both | 1085 |
| | - | - | Ap | T | 598 |
| | - | - | Km | U | 402 |
| | + | - | Ap | T | 204 |
| | + | - | Km | U | 34 |
| | + | + | Ap | T | 440 |
| | + | + | Km | U | 21 |
| 2 | + | - | Ap | T | 138 |
| | + | - | Km | U | 30 |
| | + | + | Ap | T | 336 |
| | + | + | Km | U | 31 |
| 10 | + | - | Ap | T | 67 |
| | + | - | Km | U | 3 |
| | + | + | Ap | T | 222 |
| | + | + | Km | U | 2 |
| 30 | + | - | Ap | T | 23 |
| | + | - | Km | U | 0 |
| | + | + | Ap | T | 139 |
| | + | + | Km | U | 0 |
| 60 | + | - | Ap | T | 3 |
| | + | - | Km | U | 0 |
| | + | + | Ap | T | 152 |
| | + | + | Km | U | 0 |
| 90 | + | - | Ap | T | 3 |
| | + | - | Km | U | 0 |
| | + | + | Ap | T | 114 |
| | + | + | Km | U | 0 |
| 120 | + | - | Ap | T | 0 |
| | + | - | Km | U | 0 |
| | + | + | Ap | T | 113 |
| | + | + | Km | U | 0 |
| 180 | + | - | Ap | T | 0 |
| | + | - | Km | U | 0 |
| | + | + | Ap | T | 106 |
| | + | + | Km | U | 0 |
| 240 | + | - | Ap | T | 0 |
| | + | - | Km | U | 0 |
| | + | + | Ap | T | 102 |
| | + | + | Km | U | 0 |

- 104 -

Table 5

Presence or Absence of
mAbT7 in Medium

mAbT7 + mAbT7 -

Presence or Absence
of T7 Tag at OmpA Loop IV

| | | |
|-------|----------------|----------------|
| Tag + | K3 Resistant | K3 Susceptible |
| Tag - | K3 Susceptible | K3 Susceptible |

Table 6

| Time [min] | Phage | Antibody | Selection | Selected | Number of colonies |
|------------|-------|----------|-----------|----------|--------------------|
| 0 | - | - | LB | both | 1188 |
| | - | - | Ap | T | 663 |
| | - | - | Km | U | 484 |
| | + | - | Ap | T | 12 |
| | + | - | Km | U | 0 |
| 1 | + | + | Ap | T | 18 |
| | + | + | Km | U | 0 |
| 10 | + | + | Ap | T | 24 |
| | + | + | Km | U | 1 |
| 20 | + | + | Ap | T | 43 |
| | + | + | Km | U | 0 |
| 30 | + | + | Ap | T | 75 |
| | + | + | Km | U | 0 |
| 60 | + | + | Ap | T | 655 |
| | + | + | Km | U | 0 |

Table 7

| Time [min] | Phage | Antibody | Selection | Selected | Number of cells |
|------------|-------|----------|-----------|-------------|-----------------|
| 0 | - | - | amp | T7 tag-OmpA | 616 |
| | - | - | kan | wt OmpA | 75400 |
| 120 | - | - | amp | T7 tag-OmpA | 2867 |
| | - | - | kan | wt OmpA | 476000 |
| | + | - | amp | T7 tag-OmpA | 0 |
| | + | - | kan | wt OmpA | 0 |
| | + | + | amp | T7 tag-OmpA | 2845 |
| | + | + | kan | wt OmpA | 0 |

- 105 -

Table 8

| Time | K3 | mAb | Selecti | Select | Number of | Titer of K3 | MoI |
|------|----|-----|---------|--------|-----------|-------------|-----|
| -60 | - | - | Ap | T | 269 | - | - |
| | | | Km | U | 50176 | - | - |
| 0 | - | - | Ap | T | 627 | 4.00E+05 | 6 |
| | | | Km | U | 67872 | - | - |
| 15 | - | - | Ap | T | 873 | - | - |
| | | | Km | U | 73472 | - | - |
| | + | - | Ap | T | 269 | 5.91E+05 | 11 |
| | | | Km | U | 54656 | - | |
| | + | + | Ap | T | 1120 | - | 12 |
| | | | Km | U | 52864 | - | |
| 30 | - | - | Ap | T | 1832 | - | - |
| | | | Km | U | 95200 | - | - |
| | + | - | Ap | T | 224 | 6.00E+05 | 12 |
| | | | Km | U | 50736 | - | |
| | + | + | Ap | T | 1344 | - | 21 |
| | | | Km | U | 48832 | - | |
| 45 | - | - | Ap | T | 3853 | - | - |
| | | | Km | U | 134848 | - | - |
| | + | - | Ap | T | 269 | 1.00E+06 | 21 |
| | | | Km | U | 47712 | - | |
| | + | + | Ap | T | 1344 | - | 25 |
| | | | Km | U | 47264 | - | |
| 60 | - | - | Ap | T | 2061 | - | - |
| | | | Km | U | 140000 | - | - |
| | + | - | Ap | T | 313 | 1.00E+06 | 25 |
| | | | Km | U | 36736 | - | |
| | + | + | Ap | T | 3136 | - | 20 |
| | | | Km | U | 37520 | - | |
| 75 | - | - | Ap | T | 2867 | - | - |
| | | | Km | U | 353920 | - | - |
| | + | - | Ap | T | 224 | 8.00E+05 | 20 |
| | | | Km | U | 34384 | - | |
| | + | + | Ap | T | 4144 | - | 40 |
| | | | Km | U | 35840 | - | |
| 90 | - | - | Ap | T | 2912 | - | - |
| | | | Km | U | 400960 | - | - |
| | + | - | Ap | T | 224 | 1.20E+06 | 40 |
| | | | Km | U | 25200 | - | |
| | + | + | Ap | T | 4256 | - | 40 |
| | | | Km | U | 4256 | - | |

- 106 -

| | | | | | | | |
|-----|---|---|----|---|--------|----------|----|
| | | | Km | U | 25984 | | |
| 105 | - | - | Ap | T | 4256 | - | - |
| | | | Km | U | 432320 | - | - |
| | + | - | Ap | T | 179 | 2.20E+06 | |
| | | | Km | U | 19600 | | |
| | + | + | Ap | T | 3920 | | 95 |
| | | | Km | U | 19152 | | |

| | | | | | | | |
|-----|---|---|----|---|---------|----------|-------|
| 120 | - | - | Ap | T | 4816 | - | - |
| | | | Km | U | 439040 | - | - |
| | + | - | Ap | T | 134 | 1.46E+07 | |
| | | | Km | U | 1344 | | |
| | + | + | Ap | T | 4256 | | 2834 |
| | | | Km | U | 896 | | |
| 135 | - | - | Ap | T | 5152 | - | - |
| | | | Km | U | 772800 | - | - |
| | + | - | Ap | T | 123 | 1.60E+07 | |
| | | | Km | U | 10 | | |
| | + | + | Ap | T | 3920 | | 4059 |
| | | | Km | U | 22 | | |
| 150 | - | - | Ap | T | 6272 | - | - |
| | | | Km | U | 1136800 | - | - |
| | + | - | Ap | T | 33 | 2.88E+07 | |
| | | | Km | U | 0 | | |
| | + | + | Ap | T | 3136 | | 88889 |
| | | | Km | U | 10 | | |
| 165 | - | - | Ap | T | 6720 | - | - |
| | | | Km | U | 2004800 | - | - |
| | + | - | Ap | T | 0 | 6.20E+07 | |
| | | | Km | U | 0 | | |
| | + | + | Ap | T | 2800 | | 22142 |
| | | | Km | U | 0 | | |
| 180 | - | - | Ap | T | 8512 | - | - |
| | | | Km | U | 3136000 | - | - |
| | + | - | Ap | T | 0 | 5.00E+07 | |
| | | | Km | U | 0 | | |
| | + | + | Ap | T | 2576 | | 19409 |
| | | | Km | U | 0 | | |

Table 9

| Time [min] | Helper Cells [μl] | K3 | mAb | Cell type | Number of cells | Background [%] |
|---------------|-------------------------|----|-----|--------------|--------------------|-------------------|
| -60 | - | - | - | T | 21863 | |
| | | | | U | 2446080 | |
| 0 | - | - | - | T | 26476 | |
| | | | | U | 5249760 | |
| 120 | - | - | - | T | 158112 | |
| | | | | U | 52240400 | |
| | 0 | + | - | T | 1562 | |
| | | | | U | 7466 | |
| | | + | + | T | 14600 | 33.88% |
| | | | | U | 7480 | |
| | 100 | + | - | T | 256 | |
| | | | | U | 4405 | |
| | | + | + | T | 6830 | 15.80% |
| | | | | U | 1282 | |

Table

10

| Time [min] ¹ | K3 | mAb | Cell Type | Dilution of T cell culture | Number of Cells | T:U Ratio | | |
|----------------------------|----|-----|--------------|----------------------------------|--------------------|------------|-------|------------|
| -60 | - | - | U | - | 1.46E+08 | | | |
| | | | T | undiluted | 173100 | 1.19E-03 | | |
| | | | | 10 | 21500 | 1.47E-04 | | |
| | | | | 100 | 1340 | 9.18E-06 | | |
| | | | | 1000 | 106 | 7.26E-07 | | |
| | | | | 10000 | 11 | 7.53E-08 | | |
| 0 | - | - | U | - | 1.86E+08 | | | |
| | | | T | undiluted | 153200 | | | |
| | | | | 10 | 8736 | | | |
| | | | | 100 | 985 | | | |
| | | | | 1000 | 146 | | | |
| | | | | 10000 | 15 | | | |
| 120 | - | - | U | - | 5.37E+08 | | | |
| | | | T | undiluted | 256200 | | | |
| | | | | 10 | 12400 | | | |
| | | | | 100 | 1500 | | | |
| | | | | 1000 | 270 | | | |
| | | | | 10000 | 18 | | | |
| | + | - | U | - | 446000 | | | |
| | | | T | undiluted | 2318 | | | |
| | | | | 10 | 120 | | | |
| | | | | 100 | 12 | | | |
| | | | | 1000 | 0 | | | |
| | | | | 10000 | 0 | | | |
| | + | + | U | - | 396500 | Enrichment | Yield | Background |
| | | | T | undiluted | 264500 | 5.63E+02 | 100% | 60.0% |
| | | | | 10 | 14200 | 2.43E+02 | 100% | 96.5% |
| | | | | 100 | 1500 | 4.12E+02 | 100% | 99.6% |
| | | | | 1000 | 195 | 6.77E+02 | 72% | 100.0% |
| | | | | 10000 | 3 | 1.00E+02 | 17% | 100.0% |

- 109 -

Table 11

| Time [min] | Bacteriophage K3 | anti-T7 tag mAb | Number of cells |
|------------|------------------|-----------------|-----------------|
| -60 | - | - | 76000 |
| 120 | - | - | 2043000 |
| | + | - | 3513 |
| | + | + | 5020 |

BIBLIOGRAPHY

Arp, J., et al. (1996) A source of glycosylated human T-cell lymphotropic virus type 1 envelope protein: expression of gp46 by the vaccinia virus/T7 polymerase system. Journal of Virology 70: 7349-7359.

Balint, R.F. and J.W. Larrick (1993) Antibody engineering by parsimonious mutagenesis. Gene 137: 109-118

Block, C. et al. (1996) Quantitative structure-activity analysis correlating Ras/Raf interaction in vitro to Raf activation in vivo. Nature Structural Biology 3 (3): 244-251.

Boder, E.T. and K.D. Wittrup (1997) Yeast surface display for screening combinatorial polypeptide libraries. Nature Biotechnology 15: 553-557.

Boulanger and Letellier (1988) Characterization of ion channels involved in the penetration of phage T4 into E. coli cells. Journal of Biological Chemistry 263 (20): 9767-9775.

Boyd, D. (1994) Use of gene fusion to determine membrane protein topology. pp. 144-163 in White, S.H., ed. Membrane Protein Structure: Experimental Approaches. Oxford, New York.

Brown, S. (1992) Engineered iron oxide-adhesion mutants of the *Escherichia coli* phage 1 receptor. Proceedings of the National Academy of Science, USA 89: 8651-8655.

Calendar, R. (1988) ed. The Bacteriophages vols. 1 and 2. Plenum, New York.

Charbonnier, J.B. et al. (1995) Crystal structure of the complex of a catalytic antibody Fab fragment with a transition state analog: structural similarities in esterase-like catalytic antibodies. Proceedings of the National Academy of Sciences, USA 92(25): 11721-5.

- 111 -

Chavez, R.A., and Z.A. Hall (1991) The transmembrane topology of the amino terminus of the alpha subunit of the nicotine acetylcholine receptor. Journal of Biological Chemistry 266: 15532-15538.

Chu, T.H., and R. Dornburg (1995) Retroviral vector particles displaying the antigen binding site of antibody enable cell-type specific gene transfer. Journal of Virology 69: 2659-2663.

Couch, R.B. (1996) Rhinoviruses: propagation in cell culture. In Fields, B.N. et al.: 714-717

Cross, G.A. (1990) Glycolipid anchoring of plasma membrane proteins. Annual Review of Cell Biology 6: 1-39.

Fantini, J. et al. (1997) Synthetic soluble analogs of galactosylceramide (GalCer) bind to the V3 domain of HIV-1 gp120 and inhibit HIV-1-induced fusion and entry. Journal of Biological Chemistry 272: 7245-7252.

Folgori, A., R. Tafi, A. Meola, F. Felici, G. Galfre, R. Cortese, P. Monaci, and A. Nicosia (1994) A general strategy to identify mimotopes of pathological antigens using only random peptide libraries and human sera. EMBO 13: 2236-2243.

Fourel, D., C. Hikita, J.M. Bolla, S. Mizushima, and J.M. Pages (1990) Characterization of OmpF domains involved in *Escherichia coli* K-12 sensitivity to colicins A and N. Journal of Bacteriology 172: 3675-3680.

Francisco, J.A., et al. (1993) Specific adhesion and hydrolysis of cellulose by intact *Escherichia coli* expressing surface anchored cellulase or cellulose binding domains. Biotechnology 11: 491-495.

Freudl, R. (1989) Insertion of peptides into cell-surface-exposed areas of the *Escherichia coli* OmpA protein does not interfere with export and membrane assembly. Gene 82: 229-236.

Georgiou, G. et al. (1997) Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines. *Nature Biotechnology* 15: 29-34.

Hibi, M. et al. (1990) Molecular cloning and expression of an IL-6 signal transducer, gp130. Cell 63: 1149-1157.

Janda, K.D. (1994) Tagged versus untagged libraries: methods for the generation and screening of combinatorial chemical libraries. Proceedings of the National Academy of Science, USA: 10779-10785.

Lis, H., and N. Sharon (1993) Protein glycosylation: Structural and functional aspects. European Journal of Biochemistry 218: 1-27

Marks, J.D., H.R. Hoogenbom, T.P. Bonnert, J. McCafferty, A.D. Griffiths, and G. Winter (1991) By-passing immunization: human antibodies from V-gene libraries displayed on phage. Journal of Molecular Biology 222: 581-597.

Mollenkopf, H.-J. et al. (1996) Conversion of bacterial gene products to secretion-competent fusion proteins. *Biotechnique* 21: 854-860.

Moran, P., and I.W. Caras (1991) Fusion of sequence elements from non-anchored proteins to generate a fully functional signal for glycosphosphatidylinositol membrane anchor attachment. Journal of Cell Biology 115: 1595-1600.

Morona, R., C. Kramer, and U. Henning (1985) Bacteriophage receptor area of outer membrane protein ompA of *Escherichia coli* K-12. Journal of Bacteriology 164: 539-543.

Olyhoek, A.J. et al. (1991) Cloning and expression in *E. coli* of *opc*, the gene for an unusual Class 5 outer membrane protein

- 113 -

from *Neisseria meningitidis*. Microbial Pathogenesis 11: 249-257.

Onu, M., G. Raab, K. Lau, J.A. Abraham, and M. Klagsbrun (1994) Purification and Characterization of Transmembrane Forms of Heparin-binding EGF-like Growth Factor. Journal of Biological Chemistry 269(49): 31315-21.

Pattus, F., D. Massotte, H.U. Wilmsen, J. Lakey, D. Tsernoglou, A. Tucker, and M.W. Parker (1990) Colicins: prokaryotic killer-pores. Experientia 46: 180 192.

Phizicky, E.M. and S. Fields (1995) Protein-protein interactions: methods for detection and analysis. Microbiological Reviews 59 (1): 94-123.

Singer, S.J. (1990) The structure and insertion of integral proteins in membranes. Annual Review of Cell Biology 6: 247-296.

Staunton, D.E., M.L. Dustin, H.P. Erickson, and T.A. Springer (1990) The arrangement of the immunoglobulin-like domains of ICAM-1 and the binding sites for LFA-1 and rhinovirus. Cell: 61: 243-354.

Türeci, Ö., U. Sahin, et al. (1997) Serological analysis of human tumor antigens: molecular definition and implications. Mol.Med.Today 3: 342-349.

Tyler, K.L. and B.N. Fields (1996) Pathogenesis of viral infections. In Fields, B.N. et al.: 173-218.

Yasukawa, K. et al. (1990) Purification and characterization of soluble human IL-6 receptor expressed in CHO cells. Journal of Biochemistry 108: 673-676.

Zeilestra-Ryalls, J. et al. (1993) Sequence analysis and phenotypic characterization of groEL mutations that block

- 114 -

lambda and T4 Bacteriophage growth. Journal of Bacteriology
175 (4): 1134-43.

CLAIMS

1.: A method for identifying and isolating at least one of a polypeptide of interest, a target compound or a gene of said polypeptide of interest or a gene of said target compound, wherein at least one of said polypeptide of interest and said target compound is unknown comprising the steps

- a) constructing a cell population by engineering it to express a membrane polypeptide which has said polypeptide of interest in its extra- or intracellular domain, said extra- or intracellular domain being exhibited in one or more copies on each cell's surface directly or indirectly interacting with a selection site specific for a selection agent, such that upon binding of said target compound to said polypeptide of interest said selection site of cells which exhibit such a binding of said target compound to said polypeptide of interest is modified in a manner that said selection agent may be used for selecting those cells which exhibit said binding of said target compound to said polypeptide of interest from cells which do not exhibit said binding,
- b) exposing said cell population to said target compound or a population of said target compounds for a period of time sufficient for said target compound to bind to said polypeptide of interest,
- c) exposing said cell population simultaneously with or after step b) to said selection agent for a period of time sufficient for said agent to interact with said selection site,
- d) selecting for clones which show said binding of said target compound to said polypeptide of interest ,
- e) optionally allowing said selected clones to multiply and
- f) isolating said polypeptide of interest and/or said target compound and/or said gene of said polypeptide of interest and/or said gene of said target compound.

- 116 -

2.: The method according to claim 1, wherein said interaction between said extra- or intracellular domain with said polypeptide of interest and said selection site is a sterical interaction, especially a sterical interaction which does not allow said selection agent to interact with said selection site after binding of said target compound to said polypeptide of interest.

3.: The method according to claim 1 or 2, wherein said selection site is immediately adjacent to the said extra- or intracellular domain with said polypeptide of interest.

4.: The method according to any one of claims 1 to 3, wherein said selection agent is a toxic agent.

5.: The method according to any one of claims 1 to 4, wherein said selection site is an adhesion domain for said selection agent.

6.: The method according to any one of claims 1 to 5, wherein said membrane polypeptide is selected from glycolipid-anchored polypeptides, α -helical polytopic receptors, bitopic integral membrane polypeptides, polytopic integral membrane polypeptides or membrane-bound isoforms of a secreted polypeptide.

7.: The method according to any one of claims 1 to 6, wherein said polypeptide of interest is identified and isolated.

8.: The method according to any one of claims 1 to 7, wherein said extra- or intracellular domain of said membrane polypeptide contains a highly variable region which is, upon binding to said target compound, identified and isolated as said polypeptide of interest or a section thereof.

9.: The method according to claim 8, wherein said highly variable domain comprises polypeptides encoded by a cDNA library or sections thereof.

- 117 -

10.: The method according to any one of claims 1 to 9, wherein said membrane polypeptide is heterologous for the cells.

11.: The method according to any one of claims 1 to 10, wherein said membrane polypeptide is genetically engineered to contain said extra- or intracellular domain and said selection site residing adjacent to each other on the cellular surface.

12.: The method according to any one of claims 1 to 11, wherein said target compound is composed of a mixture of compounds.

13.: The method according to any one of claims 1 to 12, characterised in that it is carried out in the form of a high-throughput screening assay wherein said cell population is exposed to a multiplicity of target compounds from which an unknown binding partner for a polypeptide of interest is to be identified and isolated.

14.: The method according to any one of claims 1 to 13, wherein said target compound is secreted by the cells expressing said membrane polypeptide.

15.: The method according to any one of claims 1 to 14, wherein said selection agent is a pathogenic agent.

16.: The method according to any one of claims 1 to 15, wherein said selection agent is a virus, preferably a bacteriophage.

17.: The method according to any one of claims 15 to 16, wherein said pathogenic agent is a modified pathogenic agent with a different effect on the cells than the corresponding naturally occurring pathogenic agent, preferably a stronger effect than the naturally occurring pathogenic agent.

18.: The method according to any one of claims 1 to 17, wherein steps b) through e) are repeated at least once, preferably comprising an increase in the selection pressure in the repeated steps.

- 118 -

19.: The method according to claim 18, wherein steps b) through e) are repeated until said cell population comprises a desired number of distinct clones.

20.: The method according to claim 18 or 19, wherein new sequence variations in said variable regions of clones selected in step d) are introduced before steps b) to e) are repeated.

21.: The method according to any one of claims 1 to 20, wherein said membrane protein exhibits an additional domain on the surface of the cell, said additional domain being adjacent to both, the variable domain and said selection site and said additional domain also interacts, preferably by binding, with said target compound.

22.: The method according to any one of claims 1 to 21, wherein said polypeptide of interest bind to said target compound and affect its function, comprising screening polypeptides exhibited by said clones selected in step d) and amplified in step e) by their ability to affect the function of said target compound.

23.: The method according to any one of claims 1 to 22, wherein said polypeptide of interest exhibits an enzymatic activity, comprising a step wherein said cell population is placed in a nutrient medium in which each cell's supply of a vital compound depends upon that cell's variable region acting enzymatically upon a compound structurally similar, but not identical to said target compound.

24.: The method according to any one of claims 1 to 23, wherein said membrane polypeptide is an α -helical polytopic receptor, a glycolipid-anchored protein, a glycoprotein or a proteoglycan.

25.: Cell population for use in a method according to any one of claims 1 to 24 comprising a library of cells, each cell exhibiting

a) a membrane polypeptide which has said polypeptide of interest in its extra- or intracellular domain and

- 119 -

b) a selection site being specific for a selection agent, such that upon binding of a target compound to the polypeptide of interest said selection site of cells which exhibit such a binding of said target compound to said polypeptide of interest is modified in a manner that said selection agent may be used for selecting those cells which exhibit said binding of said target compound to said polypeptide of interest from cells which do not exhibit said binding.

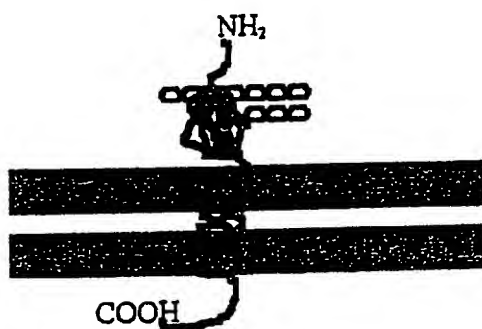
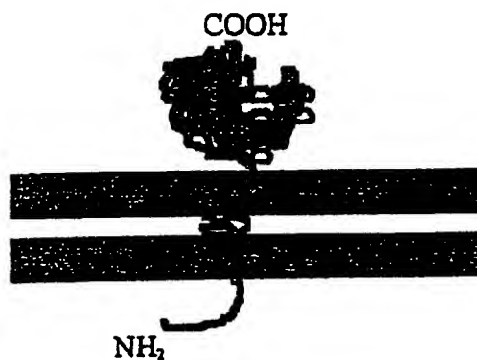
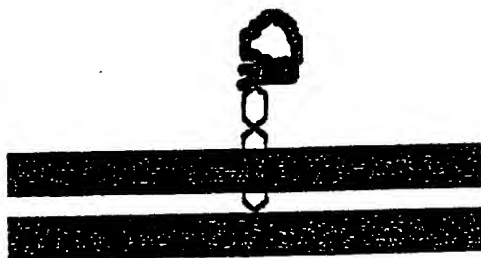
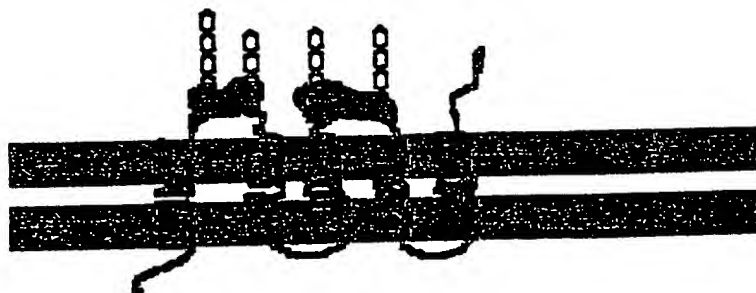
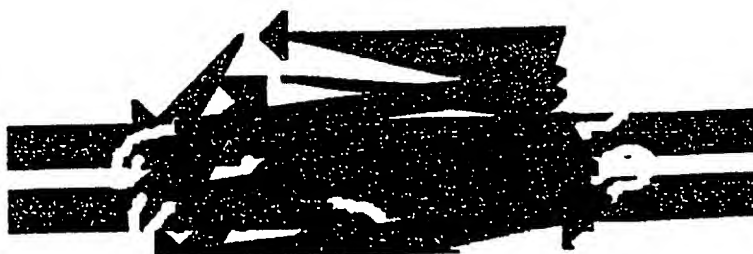
26.: Kit comprising

- a) a cell population according to claim 25,
- b) a target compound and/or a library of target compounds and
- c) one or more selection agent(s).

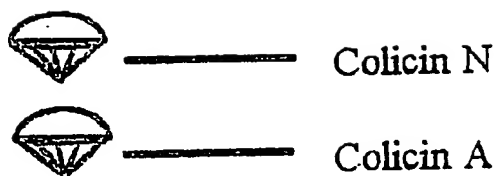
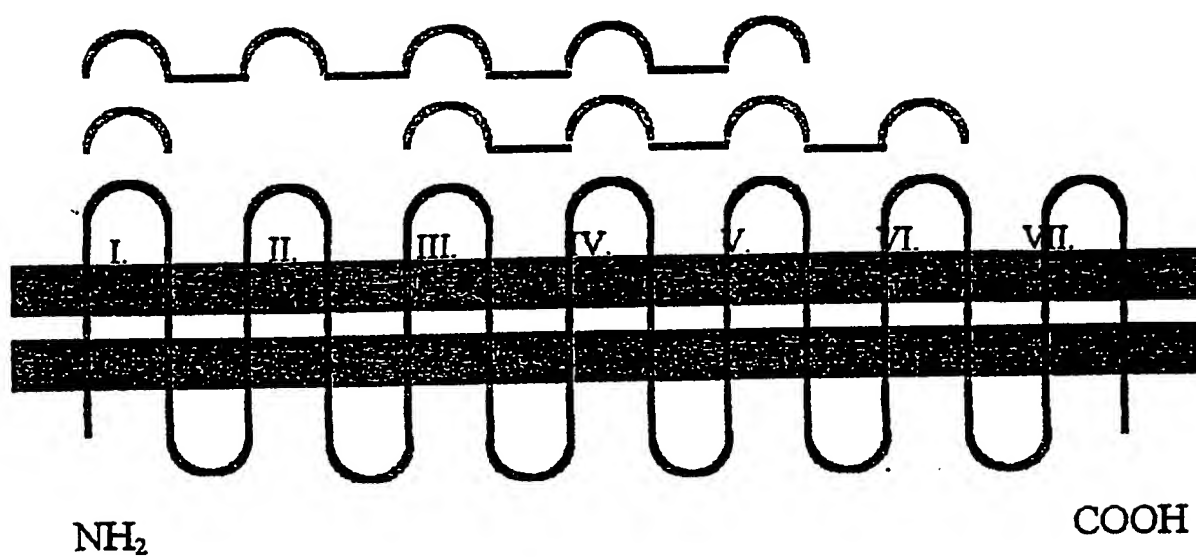
27.: A method for producing a pharmaceutical preparation, comprising

- identifying and isolating a novel pharmaceutically active compound according to a method according to any one of claims 1 to 24,
- adding pharmaceutically acceptable auxiliary components and
- completing said pharmaceutical preparation

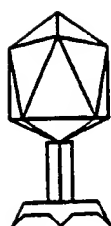
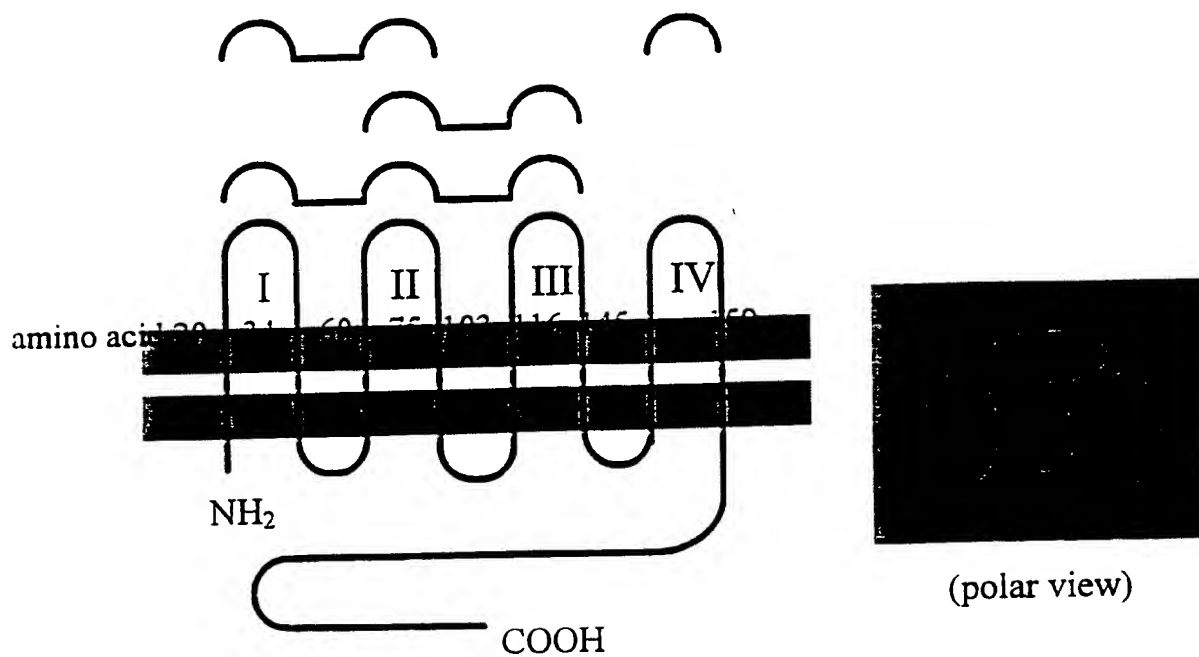
1/20

Figure 1**Type I bitopic polypeptides****Type II bitopic polypeptides****Glycolipid anchored polypeptides****Secreted polypeptides****Type III alpha-helical polytopic polypeptides****Type IV prokaryotic beta-sheet Barrels**

2/20

Figure 2

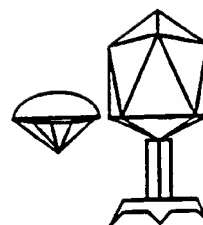
3/20

Figure 3

Tu*-46
 Tu*-60
 Tu*-24
 Tu*II-6
 Ac3
 Ox3



K3
 K4
 K5
 K3h1
 TuII*-26

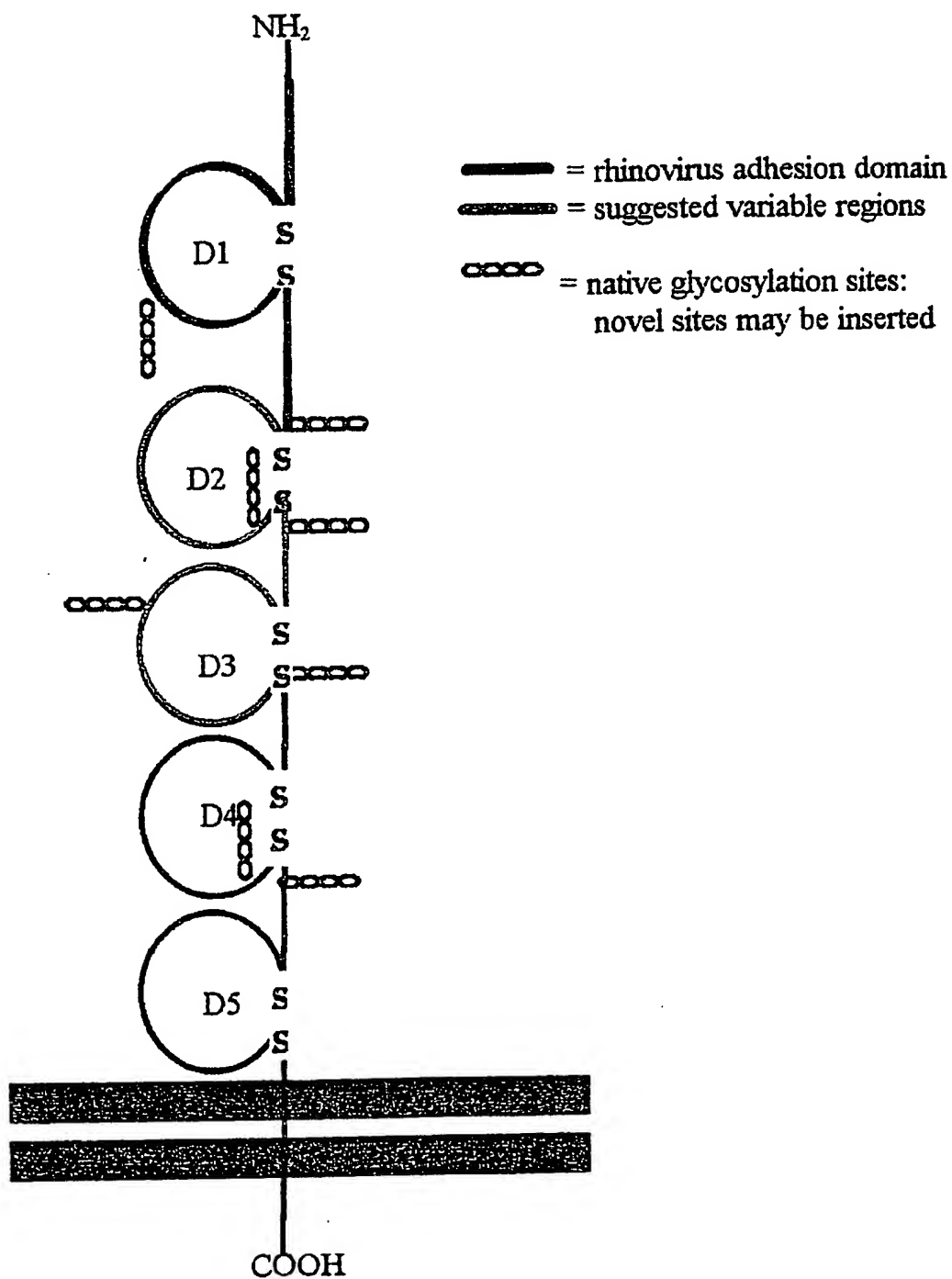


Colicin K, Ox2h20

4/20

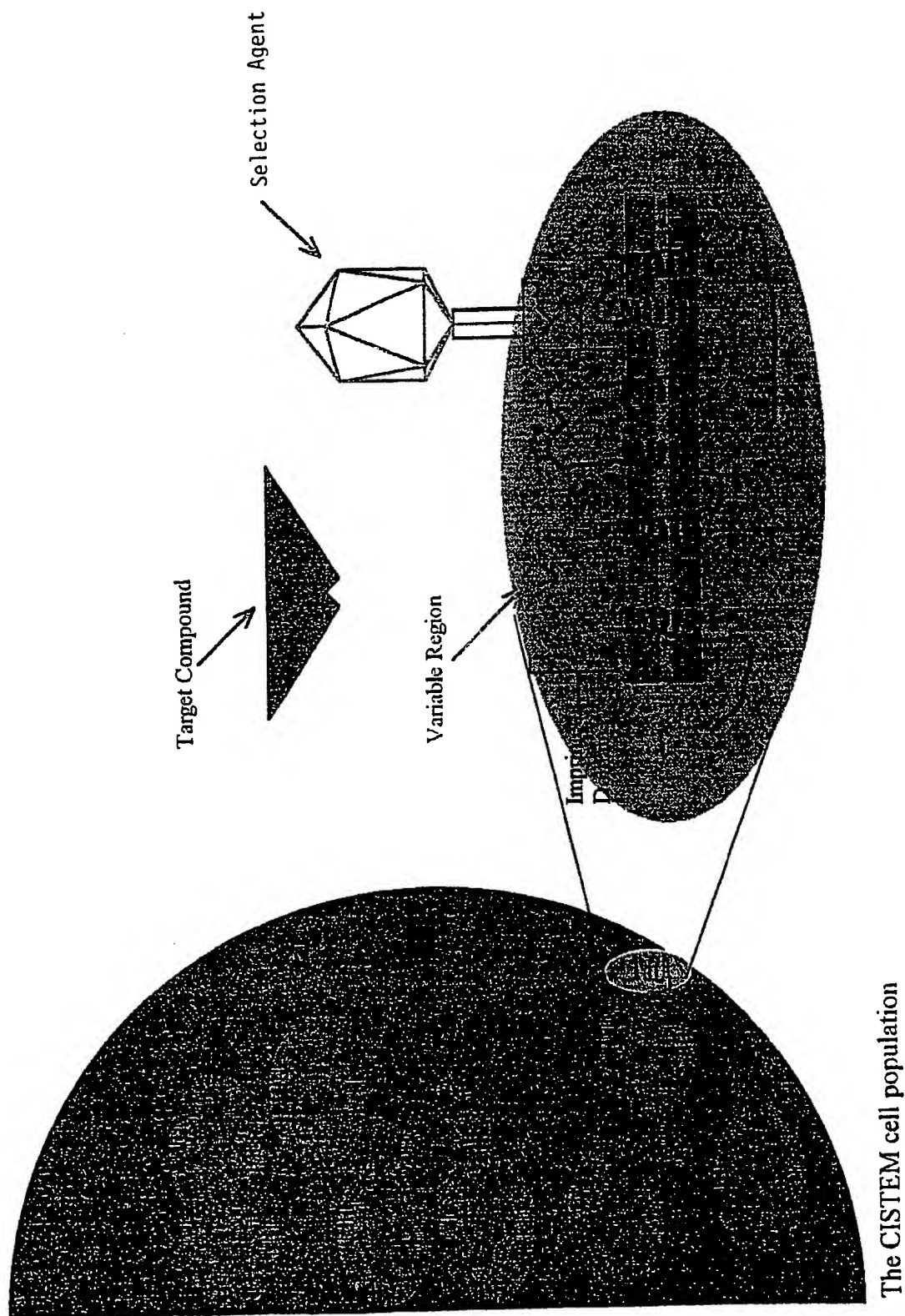
Figure 4

D1– D5 are the glycosylated protein's five homologous immunoglobulin-like domains.

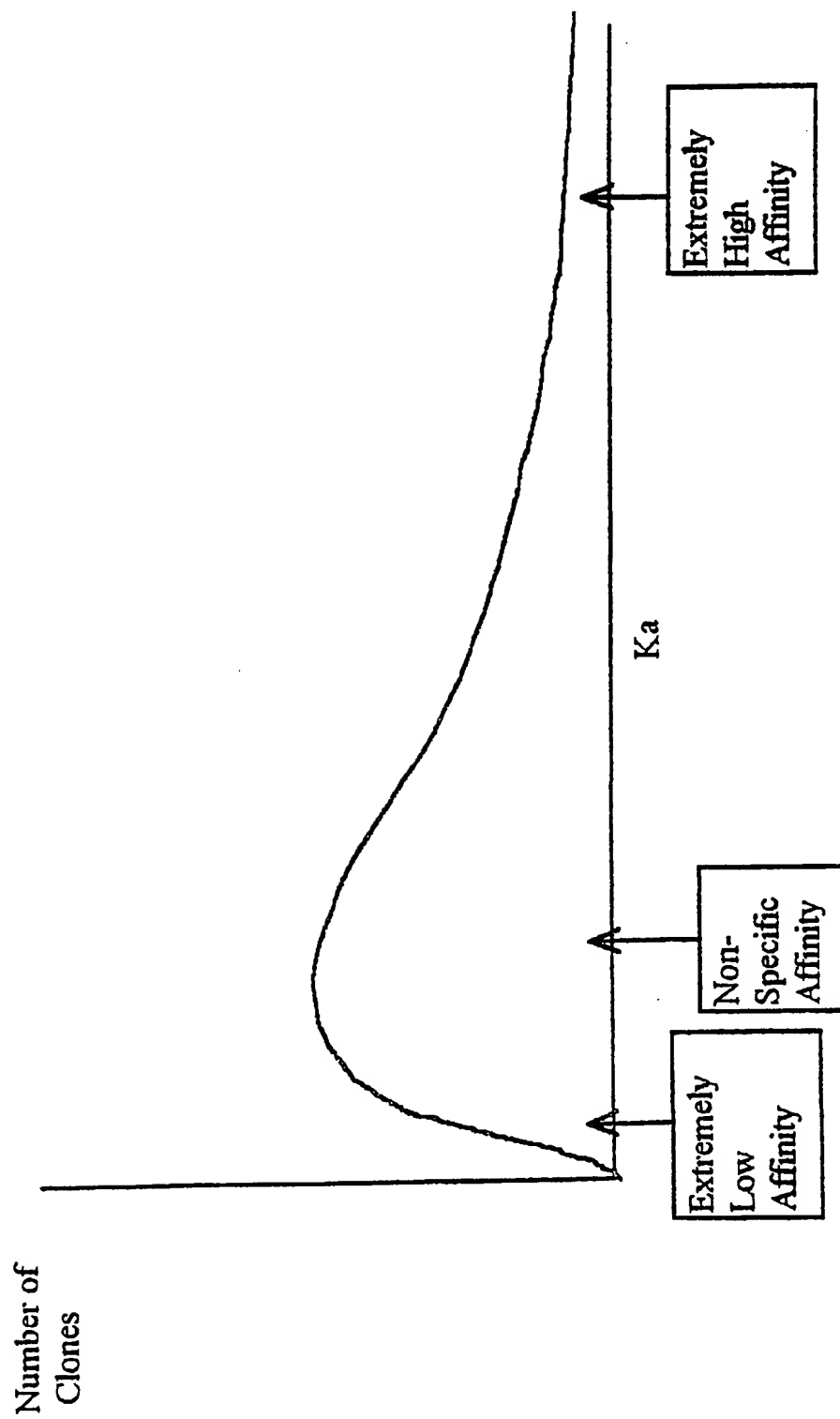


5/20

Figure 5

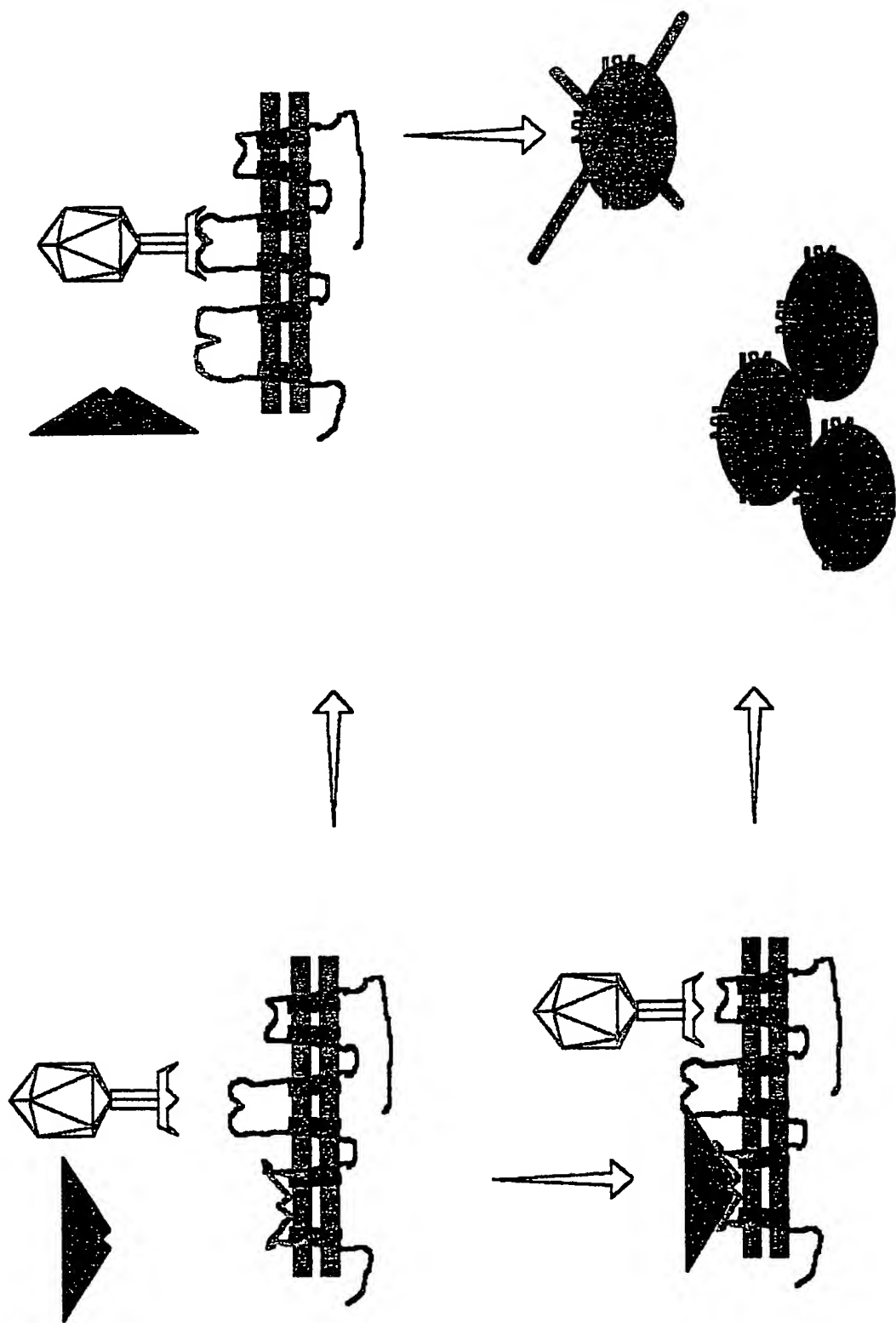


6/20

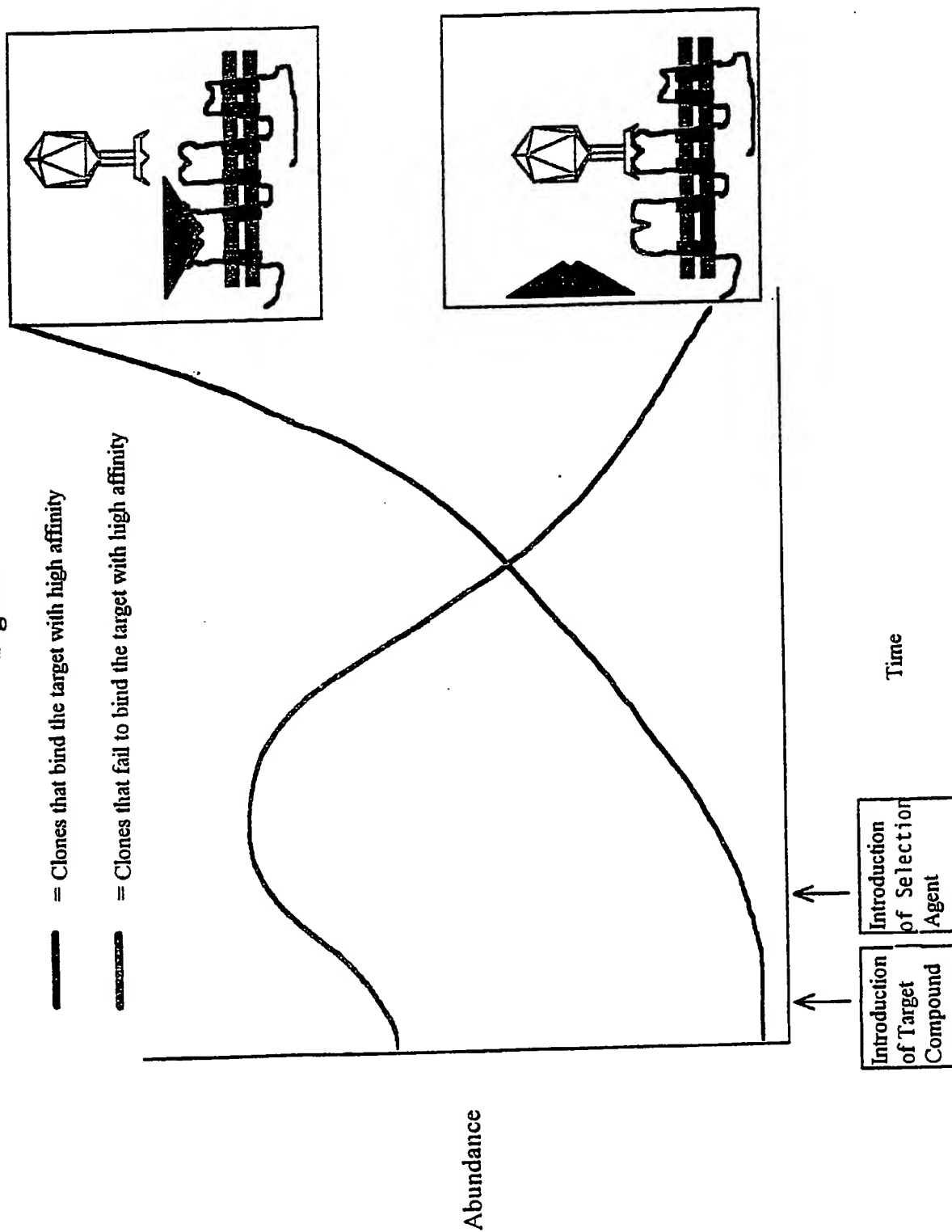
Figure 6

7/20

Figure 7

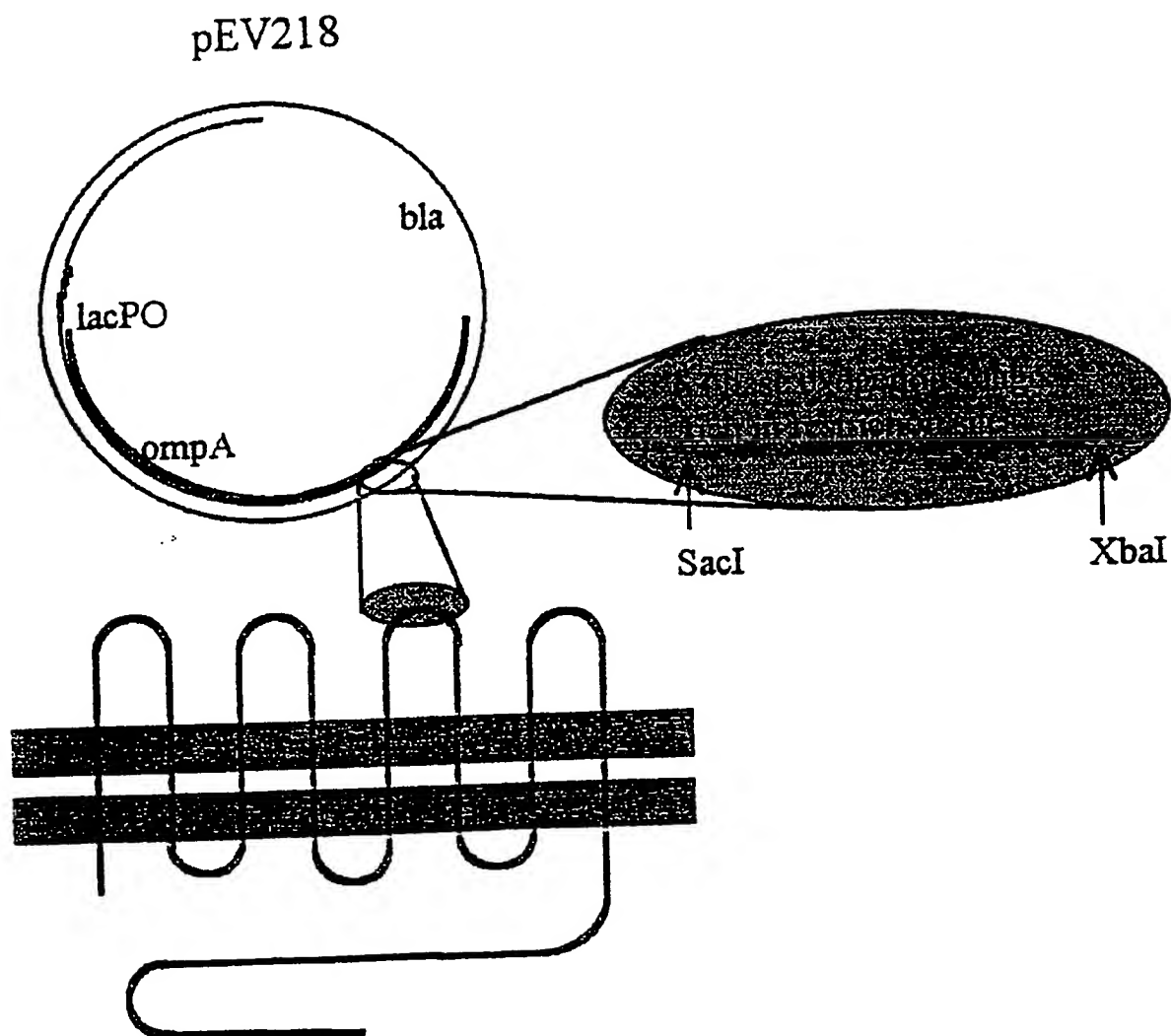


8/20

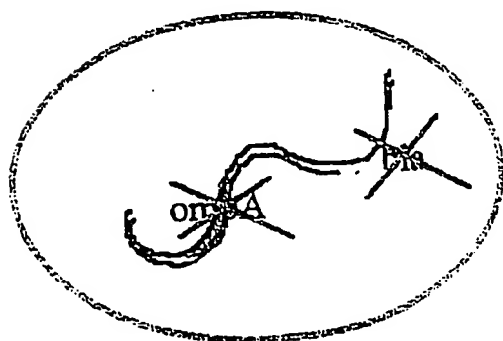
Figure 8

9/20

Figure 9

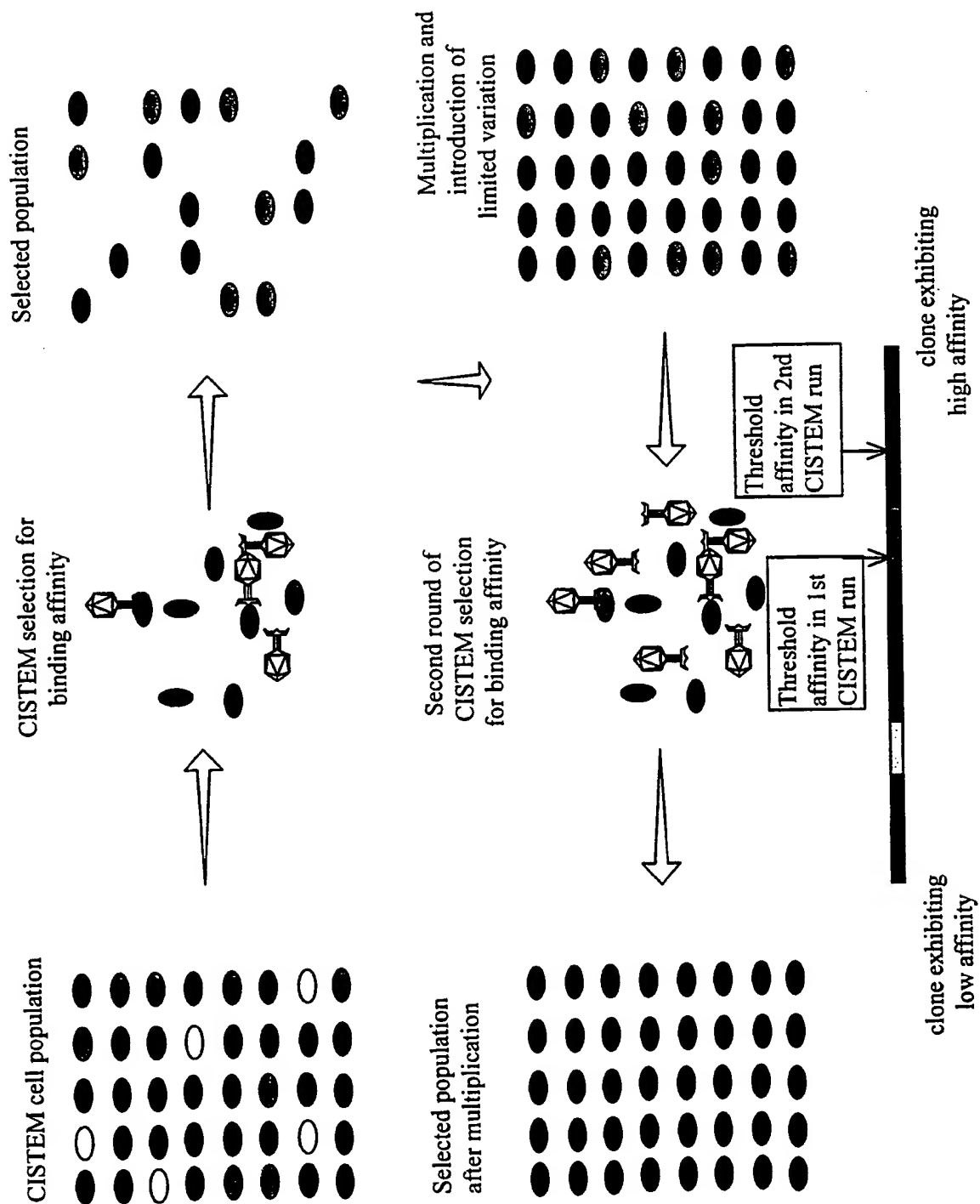


E. coli UH203



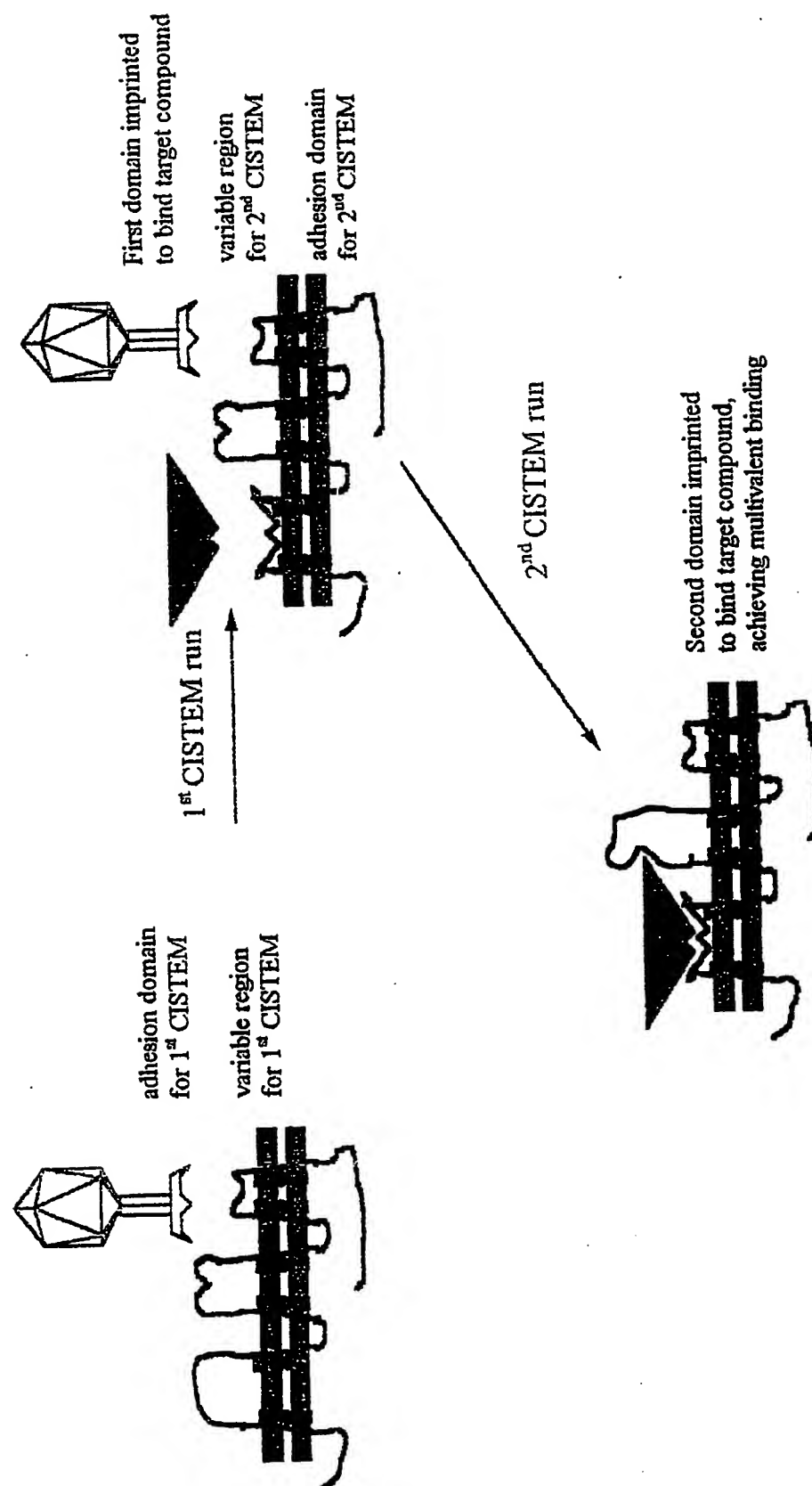
10/20

Figure 10



11/20

Figure 11



12/20

Figure 12

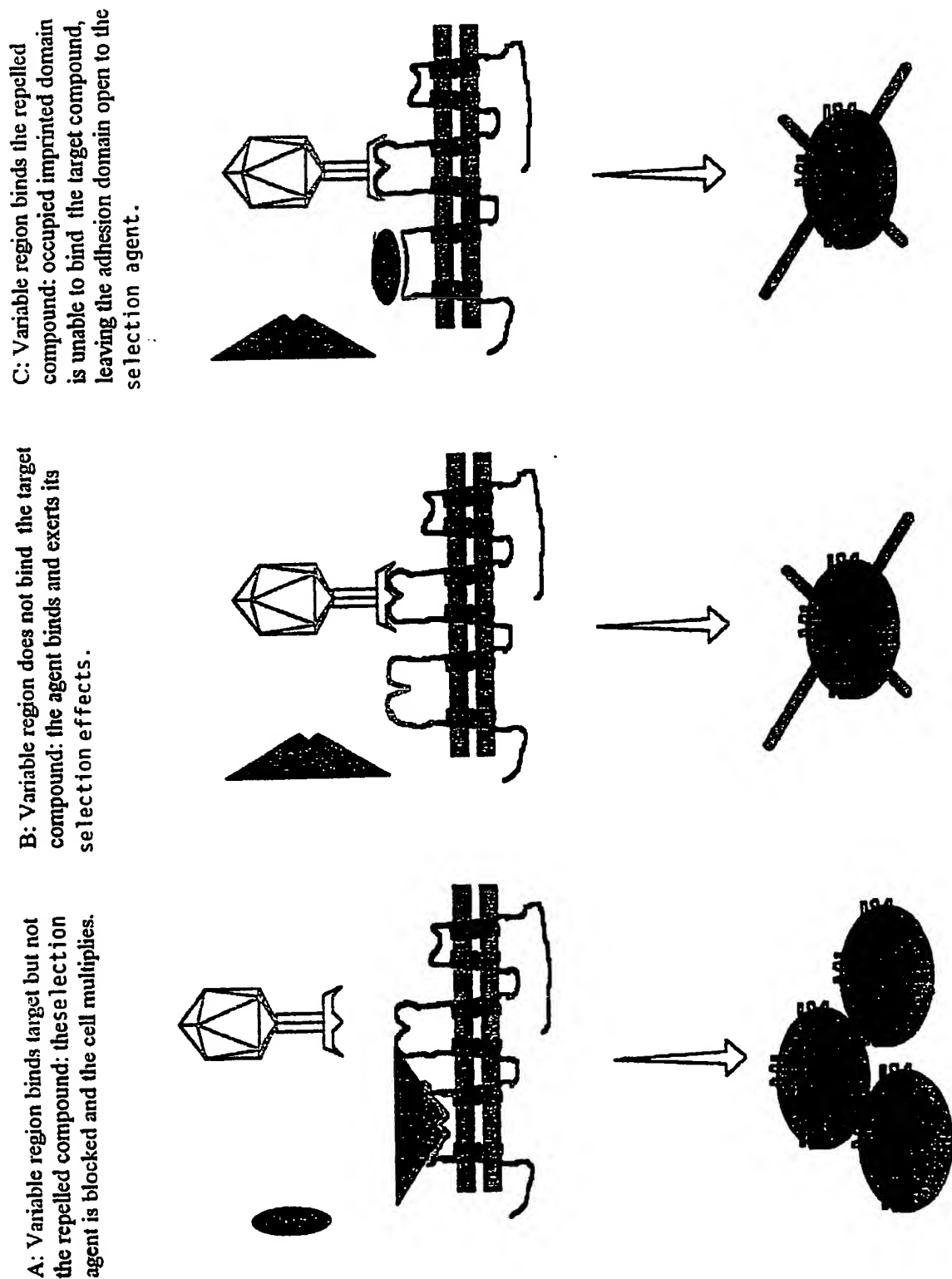
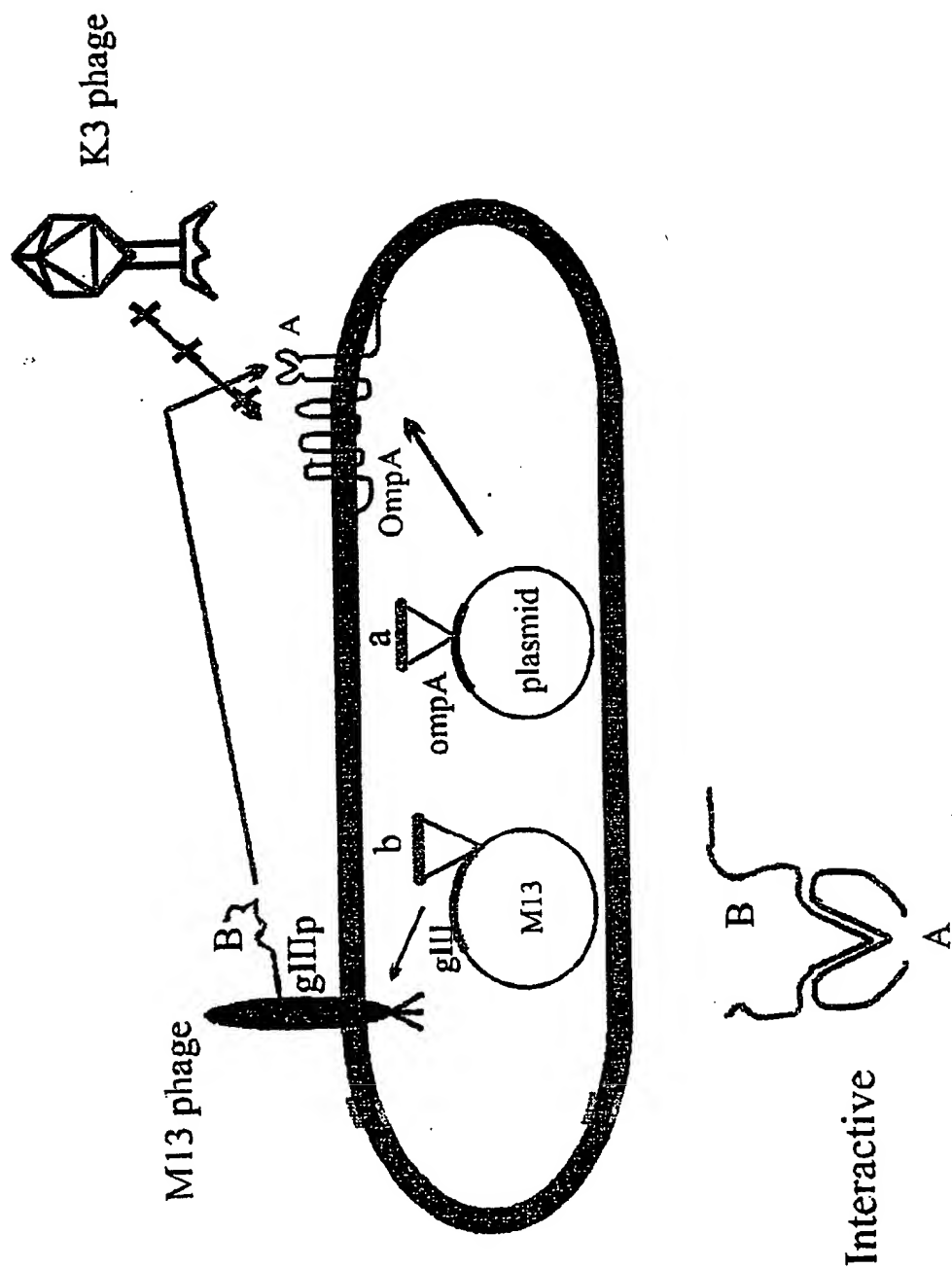


Figure 13

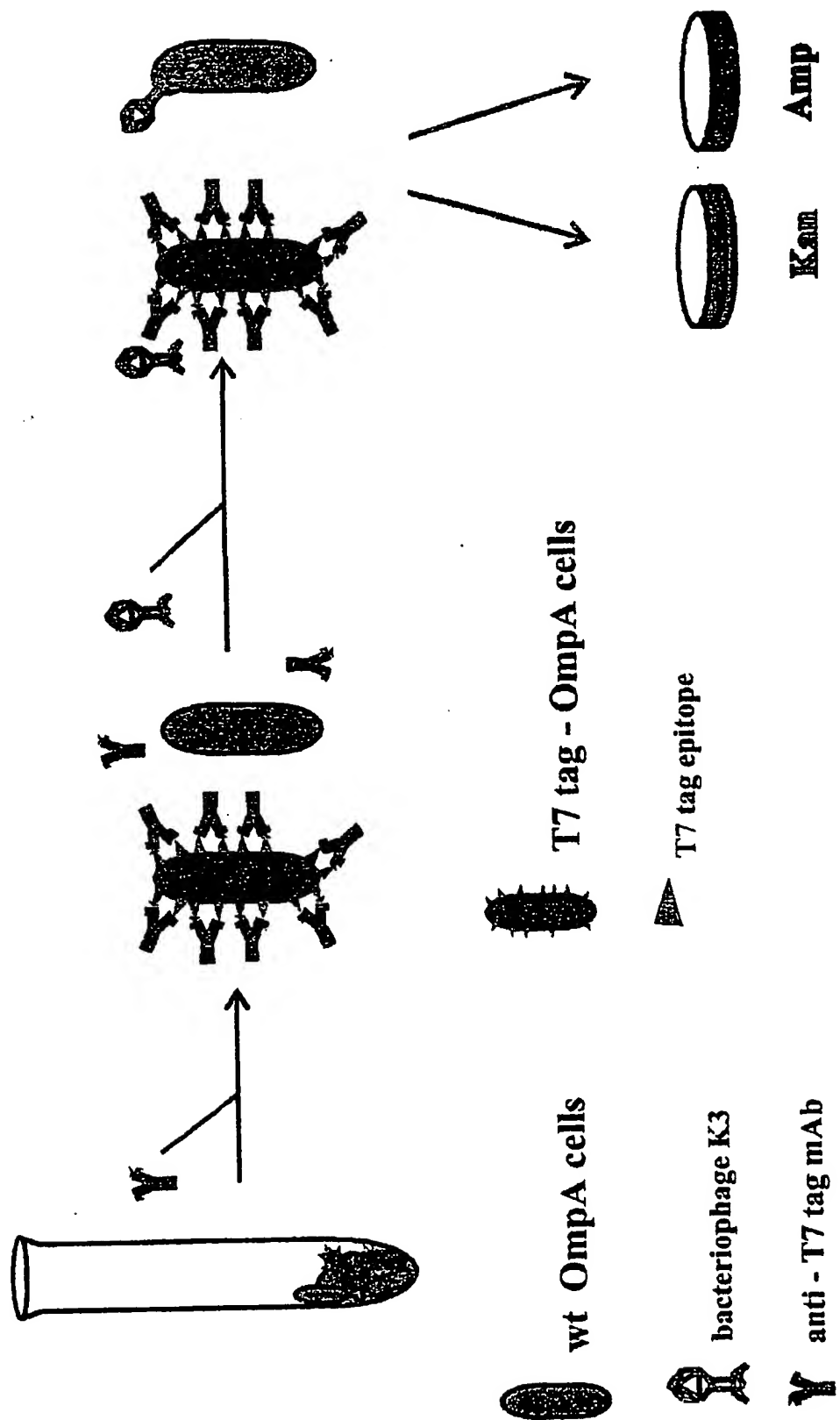
Autocrine protection by domain-domain interaction



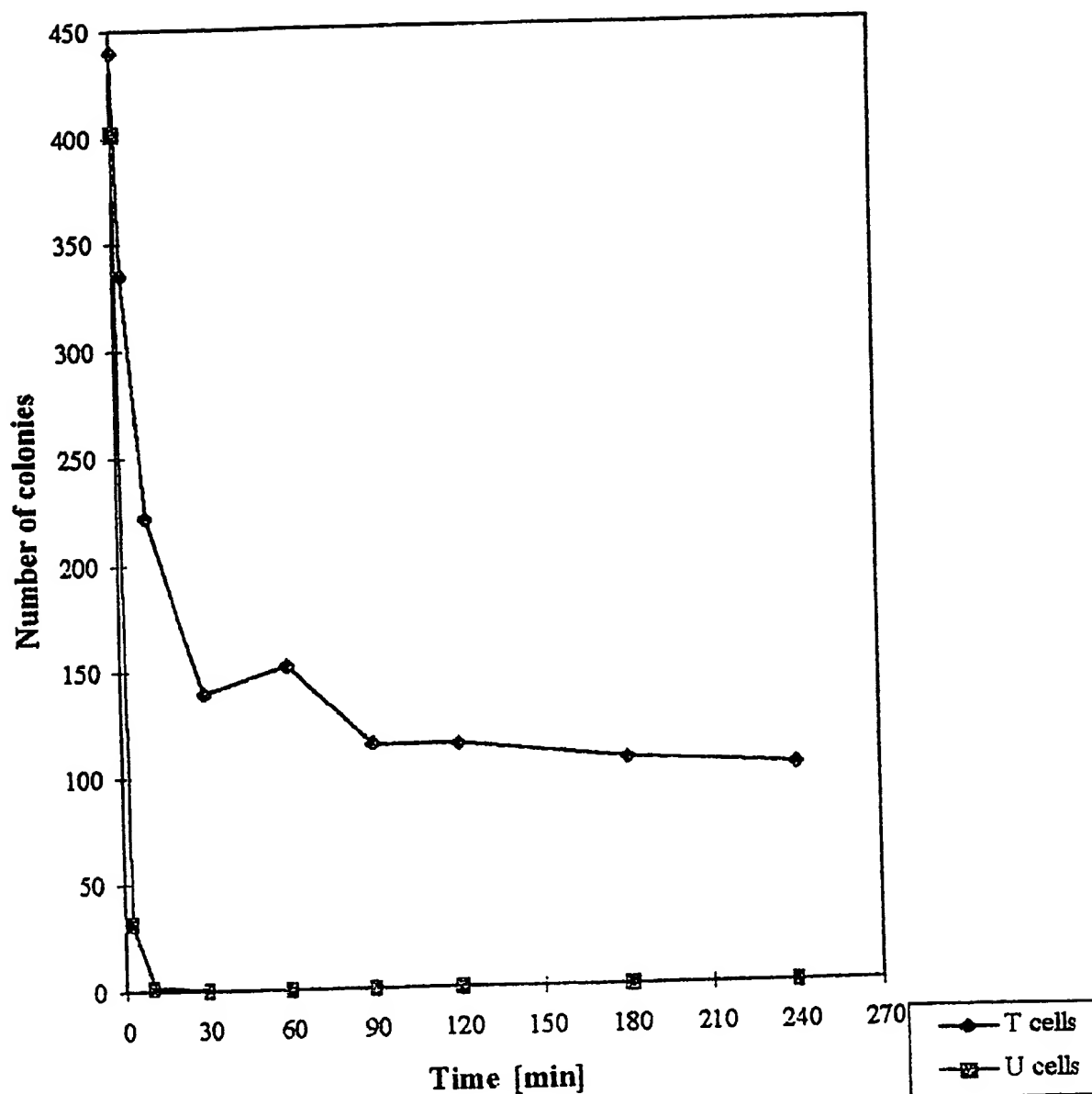
14/20

Figure 14

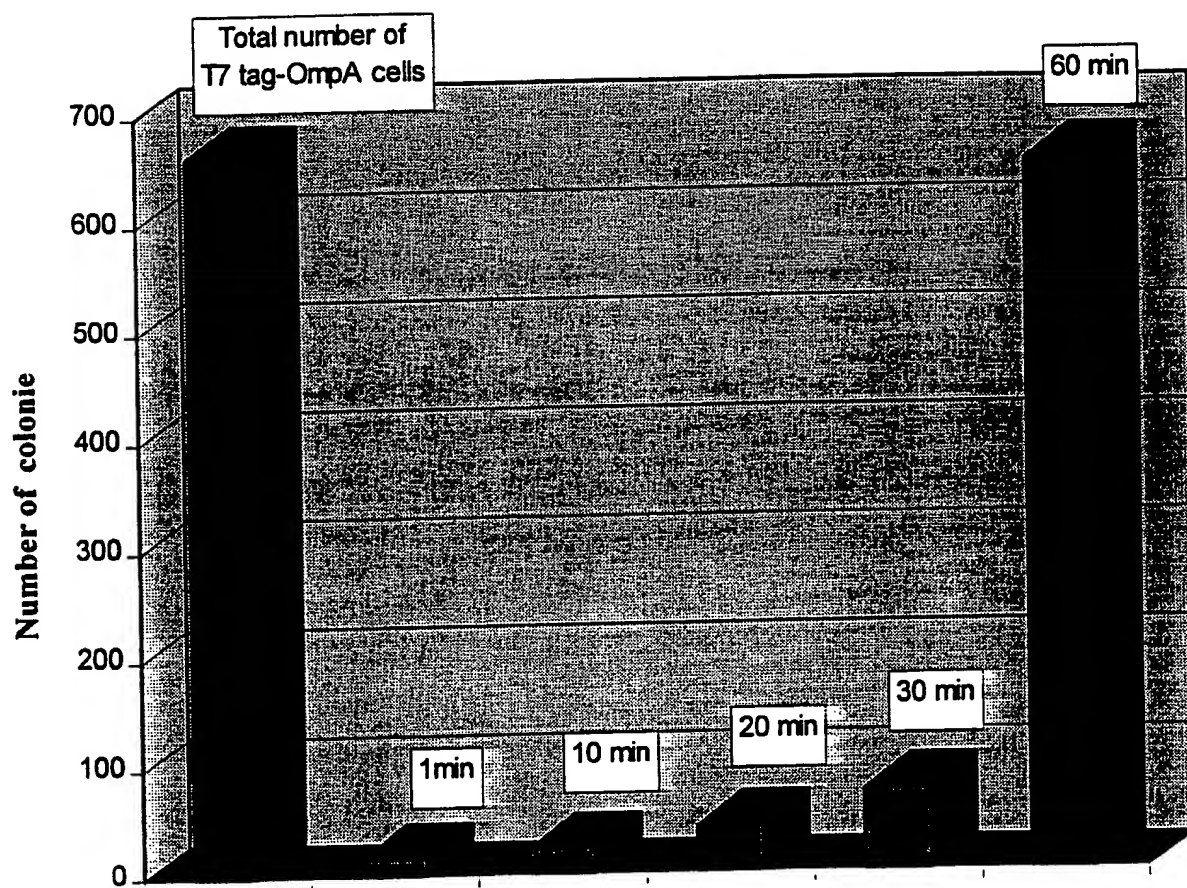
CISTEM experiment



15/20

Figure 15

16/20

Figure 16

17/20

Figure 17

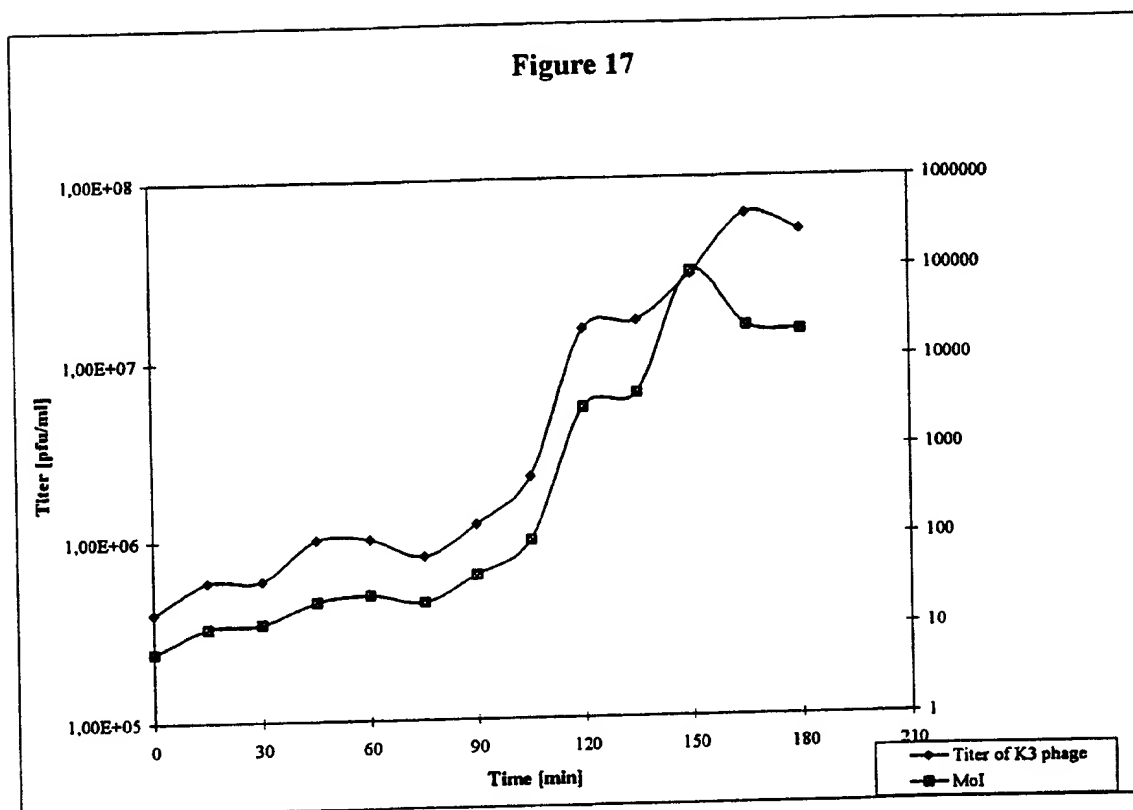
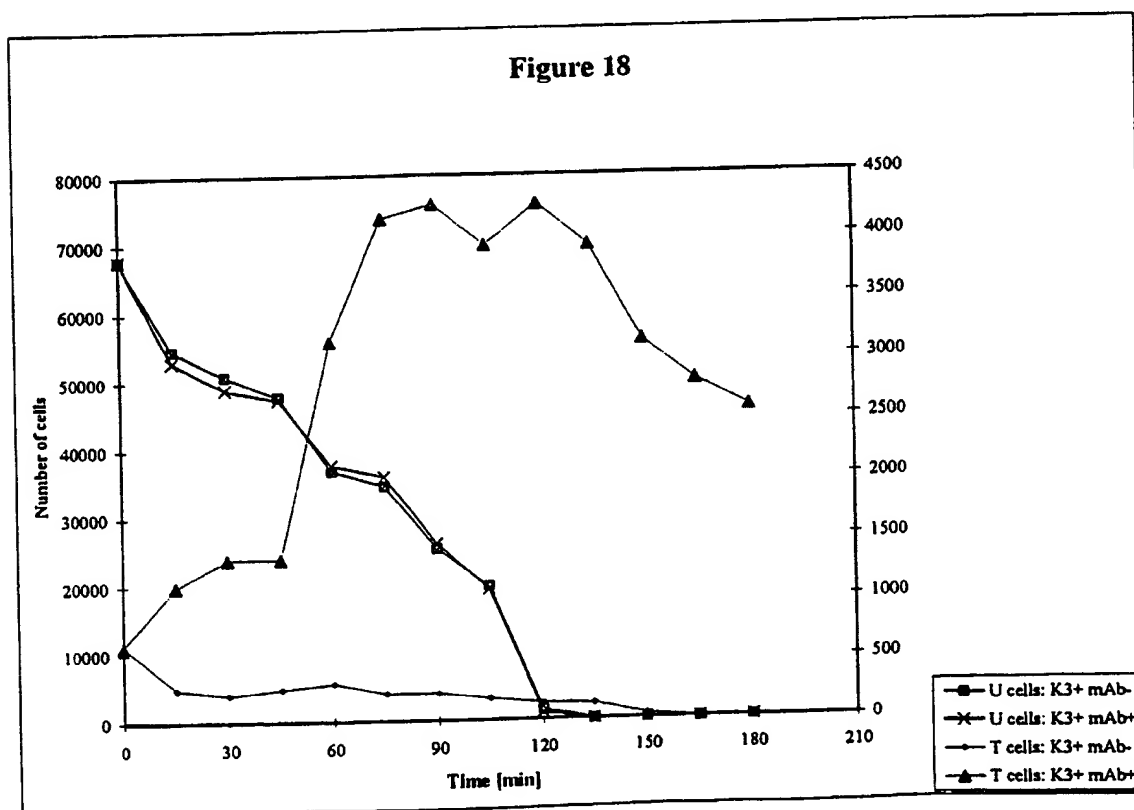
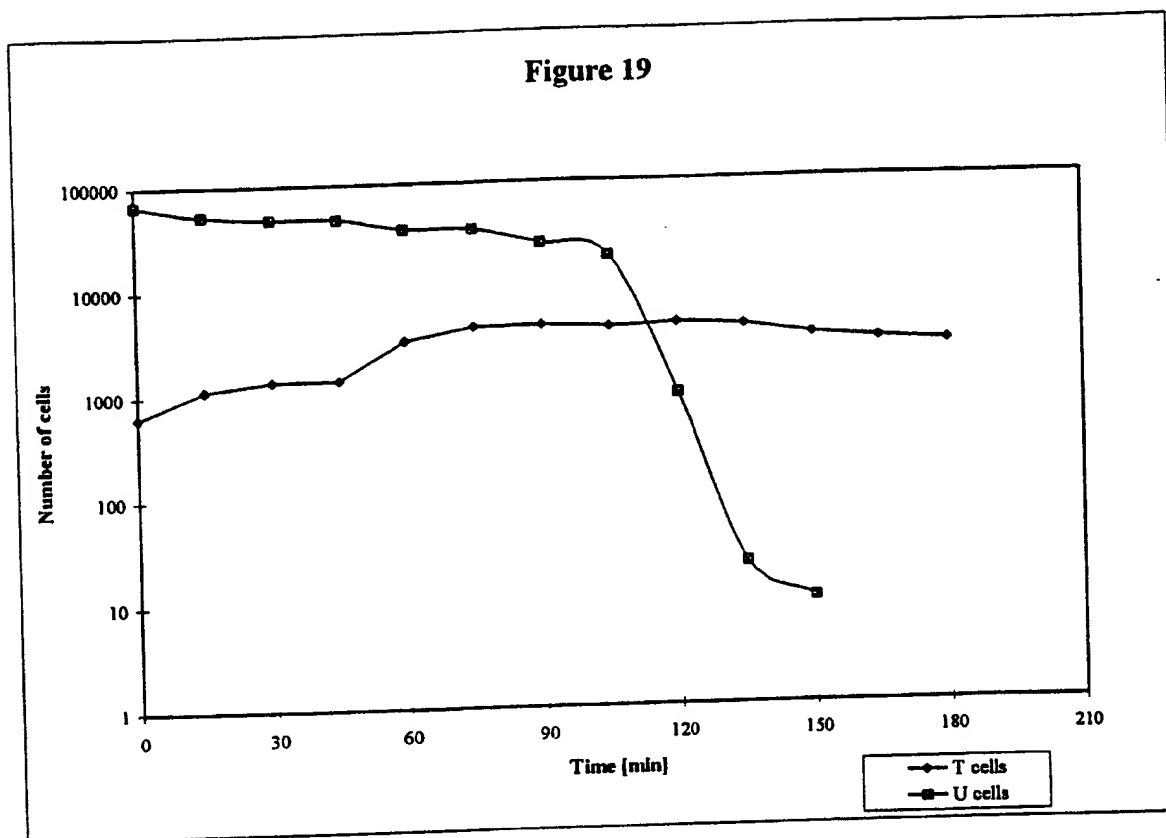


Figure 18



18/20



19/20

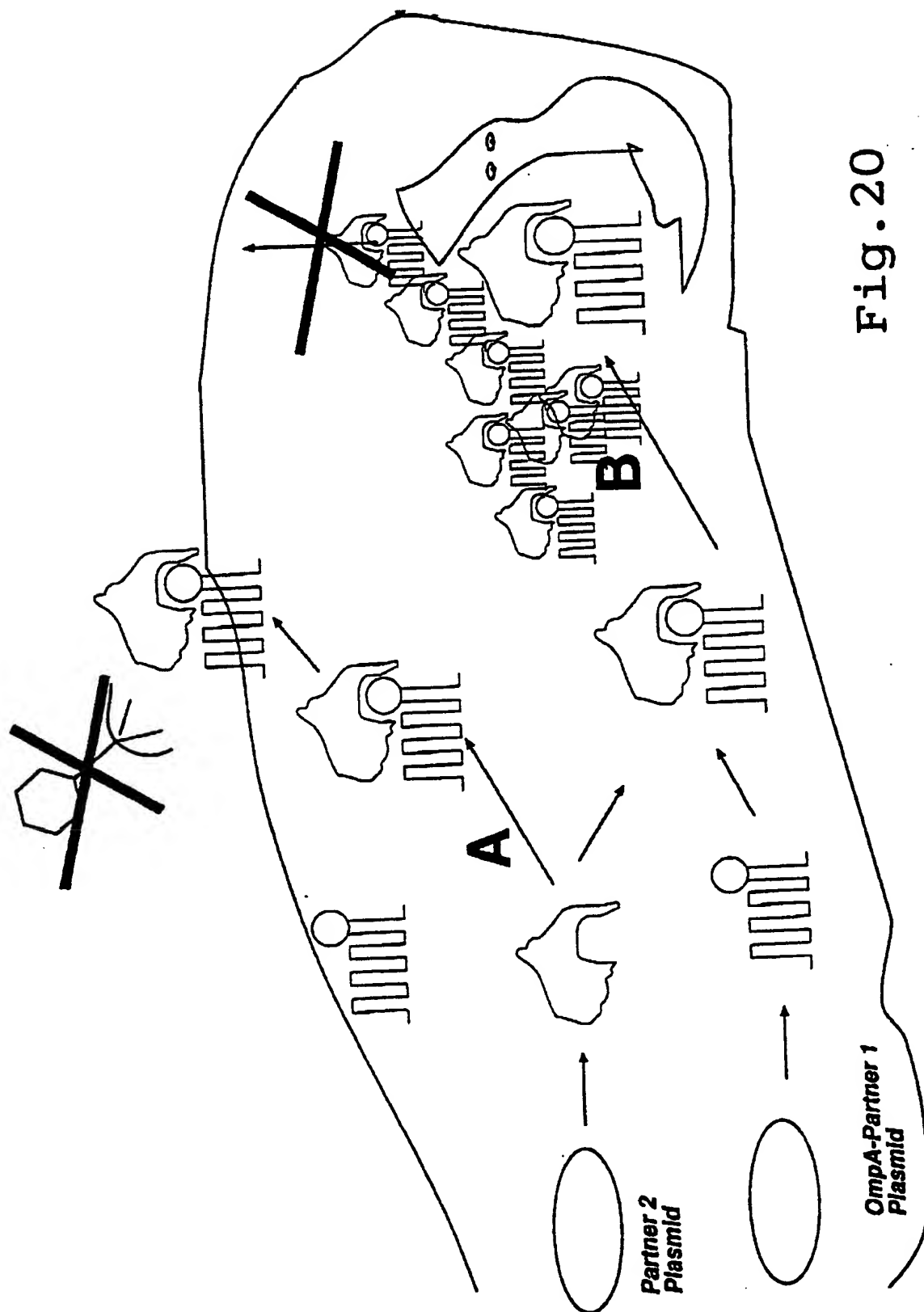
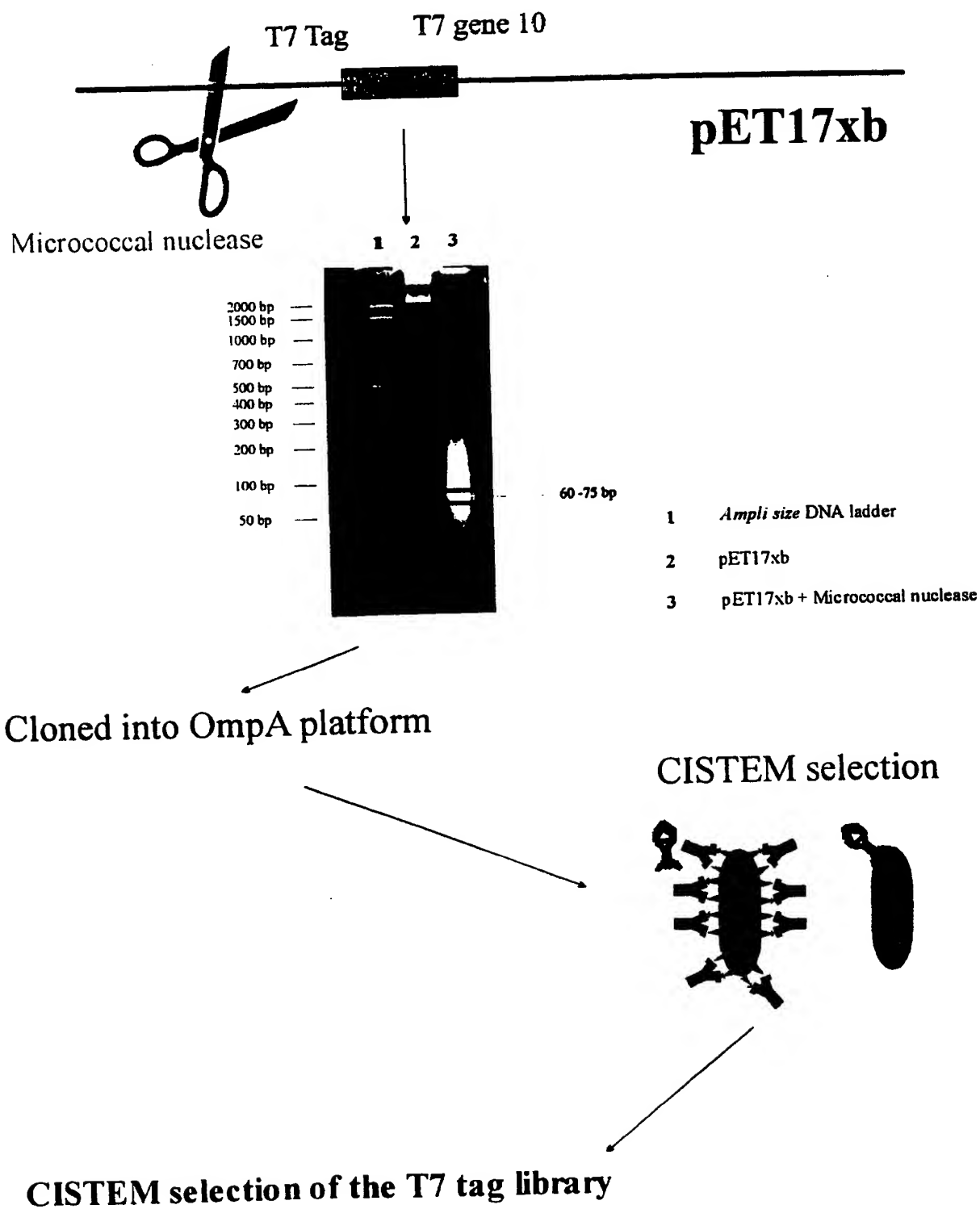


Fig. 20

20/20

Fig. 21



INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 98/07922

A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 G01N33/48 G01N33/50 G01N33/569 G01N33/68 C12N11/16

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|------------|--|-----------------------|
| X | E.T. BODER ET AL.: "Yeast surface display for screening combinatorial polypeptide libraries" NATURE BIOTECHNOLOGY, vol. 15, no. 6, 1 June 1997, pages 553-557, XP002064402 WASHINGTON DC USA cited in the application see the whole document | 1, 25 |
| A | --- | 2-24, 26, 27 |
| | --- -/-- | |

☒ Further documents are listed in the continuation of box C.

☐ Patent family members are listed in annex.

* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

27 April 1999

Date of mailing of the international search report

07/05/1999

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Van Bohemen, C

INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 98/07922

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

| Category * | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|------------|---|-----------------------|
| A | <p>M.C. KIEKE ET AL.: "Isolation of anti-T cell receptor scFv mutants by yeast surface display." PROTEIN ENGINEERING, vol. 10, no. 11, 1 November 1997, pages 1303-1310, XP002064403 OXFORD UK see the whole document -----</p> | 1-27 |